

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:31:21 ; Search time 12 Seconds

(without alignments)  
1067.438 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357

Sequence: 1 MADNFSHDLALSGSGNPNQ.....ISKLGISGIDIDLTASAYTMI 246

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 65 summaries

Database : SwissProt\_42:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1310	96.5	249	1	LEGG3_HUMAN
2	1091	80.4	295	1	P34486 canis fam11
3	1077	79.4	263	1	LEGG3_MOUSE
4	1074	79.1	261	1	P08699 rattus norv
5	1044.5	77.0	244	1	LEGG3_CRIO
6	1019	75.1	241	1	LEGG3_RABIT
7	299	22.0	353	1	LEGG9_MOUSE
8	289	21.3	355	1	LEGG9_HUMAN
9	276.5	20.4	354	1	LEGG9_RAT
10	254	18.7	323	1	LEGG4_HUMAN
11	250.5	18.5	144	1	LEGG5_PIG
12	245	18.1	301	1	LEGG6_MOUSE
13	236.5	17.4	324	1	LEGG4_RAT
14	228.5	16.8	316	1	LEGG8_HUMAN
15	218.5	16.1	316	1	LEGG8_RAT
16	215.5	15.9	323	1	LEGG4_HUMAN
17	211	15.5	135	1	LEGG7_HUMAN
18	207.5	15.3	316	1	LEGG_MOUSE
19	198	14.6	245	1	CM32_HUMAN
20	197.5	14.6	242	1	CM32_MOUSE
21	196.5	14.5	466	1	ANX7_HUMAN
22	193	14.2	1049	1	CA13_BOVIN
23	192.5	14.2	463	1	ANX7_MOUSE
24	188.5	13.9	279	1	LEGG3_MOUSE
25	188	13.9	467	1	CBPA_DICDI
26	187.5	13.8	283	1	LEGG1_HABCO
27	187	13.8	684	1	CA39_HUMAN
28	185	13.6	744	1	CA18_HUMAN
29	184	13.6	744	1	LEGG_RAT
30	183.5	13.5	743	1	LEGG_MOUSE
31	181	13.3	336	1	LEGG_HUMAN
32	180	13.3	744	1	CA18_RABIT
33	179	13.2	316	1	CC12_MOUSE

## ALIGNMENTS

RESULT 1	ID	LEGG3_HUMAN	STANDARD	PRT:	249 AA.
AC	P17931	Q16005	Q16047		
DT	01-NOV-1990	(Rel. 16, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	Galactin-3 (galactose-specific lectin 3) (MAC-2 antigen) (IGF-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (GBP 35)				
DE	(Lectin-binding protein) (Lectin L-29) (L-31) (Galactoside-binding protein) (GALBP)				
GN	LGALS3 OR MAC2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91084480; PubMed=2261464;				
RA	Robertson M.W., Albrandt K., Keller D., Liu F.-T.;				
RT	"Human IGF-binding protein: a soluble lectin exhibiting a highly conserved interspecies sequence and differential recognition of IGF glycoforms."				
RL	Biochemistry 29:8093-8100(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Cardioma;				
RX	MEDLINE=9036499; PubMed=2402511.				
RA	Cherayil B., Chaitovitz S., Wong C., Pillai S.;				
RT	"Molecular cloning of a human macrophage lectin specific for galactose."				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7324-7326(1990).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91216471; PubMed=2022338;				
RA	Oda Y., Leftler H., Sakakura Y., Kasai K.I., Baronides S.H.;				

34	179	13.2	316	1	CC13_MOUSE	P20631 caenorhabdi
35	174.5	12.9	296	1	CC01_MOUSE	P08124 caenorhabdi
36	174.5	12.9	299	1	CC04_MOUSE	P34687 caenorhabdi
37	173.5	12.8	266	1	YXWK_MOUSE	Q21184 caenorhabdi
38	173.5	12.8	512	1	ANX7_XENLA	Q92125 xenopus lae
39	173.5	12.8	754	1	CA54_MOUSE	Q28247 canis fam1
40	173	12.7	675	1	LEGG_MOUSE	O54974 mus musculu
41	173	12.7	675	1	CA39_CHICK	P35017 gallus gall
42	172.5	12.7	1464	1	CA13_MOUSE	P08121 mus musculu
43	172	12.7	301	1	CC02_MOUSE	P17656 caenorhabdi
44	171.5	12.6	464	1	S3A2_MOUSE	Q15428 homo sapien
45	171.5	12.6	1712	1	CA24_HUMAN	P08572 homo sapien
46	170	12.5	1466	1	CA13_HUMAN	P02461 homo sapien
47	169.5	12.5	302	1	CCDC_MOUSE	P17657 caenorhabdi
48	169.5	12.5	377	1	PEB1_MOUSE	Q9615 drosophila
49	169.5	12.5	505	1	ANXB_MOUSE	P50995 homo sapien
50	169	12.5	1669	1	CA14_HUMAN	P02462 homo sapien
51	168.5	12.4	1758	1	CA24_MOUSE	P17140 caenorhabdi
52	168	12.4	635	1	CA28_HUMAN	P28067 homo sapien
53	166.5	12.3	306	1	CC40_MOUSE	P34804 caenorhabdi
54	166.5	12.3	309	1	CC40_MOUSE	P17277 gallus gall
55	166.5	12.3	503	1	ANXB_MOUSE	P34777 oryctolagus
56	166.5	12.3	1262	1	CA13_MOUSE	P12105 gallus gall
57	165	12.2	210	1	CA23_MOUSE	P16252 haemochus
58	165	12.2	297	1	LEGG_MOUSE	Q09581 caenorhabdi
59	165	12.2	3124	1	CA54_MOUSE	P13944 gallus gall
60	164.5	12.1	1685	1	CA54_MOUSE	P29400 homo sapien
61	163.5	12.0	263	1	Y737_MOUSE	P83475 drosophila
62	163	12.0	864	1	EUS_MOUSE	Q09372 rattus norv
63	163	12.0	1453	1	CA11_MOUSE	P02452 homo sapien
64	163	12.0	1464	1	CA11_MOUSE	Q62203 mus musculu
65	162.5	12.0	475	1	S3A2_MOUSE	

RT "Human breast carcinoma cDNA encoding a galactoside-binding lectin  
RT homologous to mouse Mac-2 antigen.";  
RL Gene 99:279-283 (1991).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91183475; PubMed=2009535;  
RA Raz A., Cami P., Raz T., Hogan V., Mohamed A., Wolman S.R.;  
RT "Molecular cloning and chromosomal mapping of a human galactoside-  
RT binding protein.";  
RL Cancer Res. 51:2173-2178 (1991).  
RN [5]  
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.  
RX MEDLINE=93234518; PubMed=7682704;  
RA Lotz M.M., Andrews C.W., Jr., Korzelius C.A., Lee E.C.,  
RT Steele G.D., Jr., Clarke A., Mercutio A.M.;  
RT "Decreased expression of Mac-2 (carbohydrate binding protein 35) and  
RT loss of its nuclear localization are associated with the neoplastic  
RT progression of colon carcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:3466-3470 (1993).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Gastric adenocarcinoma;  
RA Kato S.;  
RT "Human galectin-3 full-length cDNA.";  
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE FROM N.A.  
RA Kadofsky M.W., Wang J.L.;  
RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RX MEDLINE=22386257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RT Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Jordan B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Manusia K., Farmer A.A., Rubin A.M., Hong L.,  
RT Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.T., Usdin J.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millar S.J.,  
RA Bosak S.A., McKwan P.J., McKernan K.J., Malek J.A., Guntaratne P.H.,  
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Murthy D.C., Sodergren E.J., Lu X., Gibbs R.A.,  
RT Fahney J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Tuchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [9]  
RP PHOSPHORYLATION.  
RX MEDLINE=94075369; PubMed=8253806;  
RA Huflejt M.E., Turck C.W., Lindstedt R., Barondes S.H., Loeffler H.;  
RT "L-29, a soluble lactose-binding lectin, is phosphorylated on serine  
RT 6 and serine 12 in vivo and by casein kinase I.";  
RL J. Biol. Chem. 268:26712-26718 (1993).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 113-249.  
RX MEDLINE=98250755; PubMed=9582341;  
RA Seetharaman J., Karigsberg A., Slaaby R., Loeffler H., Barondes S.H.,  
RT "X-ray crystal structure of the human galectin-3 carbohydrate  
RT recognition domain at 2.1-A resolution.";  
RL J. Biol. Chem. 273:13047-13052 (1998).  
CC -I- FUNCTION: Galactose-specific lectin which binds IgE.  
CC -I- SUBUNIT: Probably forms homo- or heterodimers.  
CC -I- SUBCELLULAR LOCATION: Nuclear. Cytoplasmic in adenomas and  
CC carcinomas.  
CC -I- TISSUE SPECIFICITY: A major expression is found in the colonic

CC epithelium. It is also abundant in the activated macrophages.  
CC -I- SIMILARITY: In the C-terminal section; belongs to the galectin  
CC (galactin/S-lectin) family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
CC EMBL: M57710; AAA5607.1; -;  
CC EMBL: M35368; AAA8086.1; -;  
CC EMBL: M36582; AAA36163.1; -;  
CC EMBL: M64303; -; NOT ANNOTATED\_CDS.  
CC EMBL: S59012; AAB26229.1; -;  
CC EMBL: AB006780; BAA22164.1; -;  
CC EMBL: AF031425; AAB86584.1; -;  
CC EMBL: AF031422; AAB86584.1; JOINED.  
CC EMBL: AF031423; AAB86584.1; JOINED.  
CC EMBL: AF031424; AAB86584.1; JOINED.  
CC EMBL: BC001120; AAH01120.1; -;  
CC EMBL: BC053667; AAH53667.1; -;  
CC PIR: A35820; A35820.  
CC PDB: 1A3K; 15-JUL-98.  
CC Genew: HGNC:6563; LGALS3.  
CC MIM: 153619; -;  
CC GO: GO:0005866; C:plasma membrane, TAS.  
CC GO: GO:0005530; F:lectin, TAS.  
CC InterPro: IPR008985; Cona like \_lec\_gl.  
CC InterPro: IPR01079; Galectin.  
CC Pfam: PF00337; Gal-bind\_lectin, 1.  
CC SMART: SM00276; GLECT, 1.  
CC PROSITE: PS00309; GALAPTIN, 1.  
CC Galectin; Lectin; IGF-binding protein; Repeat; Phosphorylation;  
CC Acetylation; Nuclear protein; Polymorphism; 3D-structure.  
CC INIT MET 0 1  
CC MOD\_RES 5 5  
CC MOD\_RES 11 11  
CC MOD\_RES 35 108  
CC FT  
CC REPEAT 35 43  
CC REPEAT 44 52  
CC REPEAT 53 61  
CC REPEAT 62 68  
CC REPEAT 69 77  
CC REPEAT 78 87  
CC REPEAT 88 99  
CC REPEAT 100 108  
CC DOMAIN 117 249  
CC DISULFID 172 172  
CC BINDING 180 186  
CC VARIANT 63 63  
CC FT  
CC VARIANT 97 97  
CC FT  
CC CONFLICT 32 51  
CC FT  
CC CONFLICT 87 87  
CC FT  
CC CONFLICT 231 231  
CC FT  
CC STRAND 120 120  
CC FT  
CC TURN 122 123  
CC FT  
CC TURN 128 128  
CC FT  
CC STRAND 129 137  
CC FT  
CC TURN 139 140  
CC FT  
CC STRAND 143 150  
CC FT  
CC TURN 151 152  
CC FT  
CC STRAND 153 161  
CC FT  
CC STRAND 169 176  
CC FT  
CC TURN 177 178  
CC FT  
CC STRAND 179 180

FT STRAND 184 186  
FT TURN 193 194  
FT STRAND 196 203  
FT STRAND 207 212  
FT TURN 213 214  
FT STRAND 215 221  
FT HELIX 227 229  
FT STRAND 232 237  
FT STRAND 239 249  
SQ SEQUENCE 249 AA; 26057 MW; ACBS42752D137650 CRC64;

Query Match 96.5%; Score 1310; DB 1; Length 249;  
Best Local Similarity 98.4%; Pred. No. 1.2e-74;  
Matches 245; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 2 ADNFSIHMLSSGSGNPNQGMFGAWGNOPAGAGYPGASYP-YGQAPPGAGYPPGQAPPG 60  
DB 1 ADNFSIHMLSSGSGNPNQGMFGAWGNOPAGAGYPGASYPGAGYPPGQAPPG 60  
QY 61 AVHGAAPGAPGAPAGYPPGSGGAYPSSGQSPAPGAY-ATGPGAPAGPLIYPNLP 119  
DB 61 AVHGAAPGAPGAPAGYPPGSGGAYPSSGQSPAPGAYATGPGAPAGPLIYPNLP 120  
QY 120 LPGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHF-PRPNENNRRTVTCNTKLDNNW 178  
DB 121 LPGVVPRMLITLITGVKPNANRIALDFQRGNDVAFHFPRPNENNRRTVTCNTKLDNNW 180  
QY 179 GREEROSVPPESGKPKIQLVDPDHFVAVNDAAH-LOYNHRVKKLNEISKLGISGDD 237  
DB 181 GREEROSVPPESGKPKIQLVDPDHFVAVNDAAHLOYNHRVKKLNEISKLGISGDD 240  
QY 238 LTSASYTMI 246  
DB 241 LTSASYTMI 249

RESULT 2  
LEG3 CANFA STANDARD; PRT; 295 AA.

AC P38486:  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Galectin-3 (Galectose-specific lectin 3) (MAC-2 antigen) (IGF-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (CBP 35)  
DE (Laminin-binding protein) (lectin L-29).  
GN LGALS3.

OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI\_TaxID=9615;  
RN [1]  
RC SEQUENCE OF 11-295 FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Kidney epithelium;  
RA MEDLINE=94075368; PubMed=8253805;  
RA Hermann U., Turk C.W., Atchison R.E., Huflejt M.E., Poulter L.,  
RA Gilt M.A., Burlingame A.L., Barondes S.H., Loeffler H.;  
RT "Primary structure of the soluble lactose binding lectin L-29 from  
RT rat and dog and interaction of its non-collagenous proline-,  
RT glycine-, tyrosine-rich sequence with bacterial and tissue  
RT collagenase.";  
RT J. Biol. Chem. 268:26704-26711(1993).  
RL [2]  
RP PHOSPHORYLATION;  
RX MEDLINE=94075369; PubMed=8253806;  
RX Huflejt M.E., Turk C.W., Lindstedt R., Barondes S.H., Loeffler H.;  
RT "L-29, a soluble lactose-binding lectin, is phosphorylated on serine  
RT 6 and serine 12 in vivo and by casein kinase I.";  
RL J. Biol. Chem. 268:26712-26718(1993).  
CC -!- FUNCTION: Galactose-specific lectin which binds IGF.  
CC -!- SUBUNIT: Probably forms homo- or heterodimers.  
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. May be secreted by  
CC a non-classical secretory pathway.

CC -!- PTM: The degree of phosphorylation is higher in the cytoplasmic  
CC form than in the nuclear form.  
CC -!- SIMILARITY: In the C-terminal section; belongs to the galectin  
CC (galactin/s-lectin) family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L23429; AAA16211.1; -.  
DR HSSP; P17931; 1A3K.  
DR InterPro; IPR009895; Cona-like\_lect\_g1.  
DR InterPro; IPR001079; Galectin.  
DR Pfam; PF00337; Gal-bind\_lectin; 1.  
DR SMART; SM00276; GLECT; 1.  
DR ProSite; PS00309; GALAPTIN; 1.  
DR Galectin; lectin; IGF-binding protein; Repeat; Phosphorylation;  
KW Acetylation; Nuclear protein.  
FT INTR\_MET 0 0  
FT MOD\_RES 1 1 ACETYLATION (BY CK1) (90%).  
FT MOD\_RES 5 5 PHOSPHORYLATION (BY CK1) (10%).  
FT MOD\_RES 11 11 12 X 9 AA TANDEM REPEATS OF Y-P-G-X(3) -P-  
FT DOMAIN 35 142 G-[GAT].  
FT REPEAT 35 43 1.  
FT REPEAT 44 52 2.  
FT REPEAT 53 61 3.  
FT REPEAT 62 70 4.  
FT REPEAT 71 79 5.  
FT REPEAT 80 88 6.  
FT REPEAT 89 97 7.  
FT REPEAT 98 106 8. (APPROXIMATE).  
FT REPEAT 107 114 9. (APPROXIMATE).  
FT REPEAT 115 123 10.  
FT REPEAT 124 133 11 (APPROXIMATE).  
FT REPEAT 134 142 12 (APPROXIMATE).  
FT DOMAIN 152 295 GALAPTIN.  
FT DISULFID 218 218 INTERCHAIN (BY SIMILARITY).  
FT BINDING 226 232 BETA-GALACTOSIDE (BY SIMILARITY).  
SQ SEQUENCE 295 AA; 30199 MW; 4921327063CE41C7 CRC64;

Query Match 80.4%; Score 1091; DB 1; Length 295;  
Best Local Similarity 71.2%; Pred. No. 4.6e-61;  
Matches 210; Conservative 13; Mismatches 22; Indels 50; Gaps 6;

QY 2 ADNFSIHMLSSGSGNPNQGMFGAWGNOPAGAGYPGASYP----- 42  
DB 1 ADNFSIHMLSSGSGNPNQGMFGAWGNOPAGAGYPGASYPGAGYPPGQAPPG 60  
QY 43 -----GYGQAPPGAGYPPGAGYHG--AGAYPGAPAGYPPP----- 81  
DB 61 GYGPQAPPGGYGQAPPGGYPPGQAPPGGYPGQAPPGGYPGQAPPGYPPAPYPGPTA 120  
QY 82 -----PSGPGAYSSGQSPAPGAY-ATGPGAPAGPLIYPNLPPLPGVPRMLITL 133  
DB 121 PGTOGQPSGPGAYPPPGQSPAPGAYPAAGPGIAGPLITVYDLPPLPGVPRMLITL 180  
QY 134 GTVKENANRIALDFQRGNDVAFHF-PRPNENNRRTVTCNTKLDNNWGREEROSVPPESG 192  
DB 181 GTVKENANRIALDFQRGNDVAFHFPRPNENNRRTVTCNTKLDNNWGREEROSVPPESG 240  
QY 193 KEFKIQLVDPDHFVAVNDAAH-LOYNHRVKKLNEISKLGISGDDLTSAASYTMI 246  
DB 241 KEFKIQLVDPDHFVAVNDAAHLOYNHRVKKLNEISKLGISGDDLTSAASYTMI 295

RESULT 3  
LEG3\_MOUSE STANDARD; PRT; 263 AA.

AC P16110;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Galactin-3 (Galactose-specific lectin 3) (MAC-2 antigen) (IGE-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (CBP 35) (Laminin-binding protein) (Lectin L-29) (L-34 galactoside-binding lectin).  
 DE IGAUS3.  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA/2J; TISSUE=Macrophage;  
 RA MEDLINE=90063462; PubMed=2584931;  
 RT Cherayil B.V., Wehner S.V., Pillai S.;  
 RT "The Mac-2 antigen is a galactose-specific lectin that binds IGE."; J. Exp. Med. 170:1959-1972(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=88198129; PubMed=3360772;  
 RA Jia S., Wang J.L.;  
 RT "Carbohydrate binding protein 35. Complementary DNA sequence reveals homology with proteins of the heterogeneous nuclear RNP."; J. Biol. Chem. 263:6009-6011(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=89275058; PubMed=2525069;  
 RA Raz A., Pazerini G., Carmi P.;  
 RT "Identification of the metastasis-associated, galactoside-binding lectin as a chimeric gene product with homology to an IGF-binding protein."; Cancer Res. 49:3489-3493(1989).  
 RN [4]  
 RP SEQUENCE OF 158-162; 165-174 AND 213-225.  
 RC MEDLINE=90236991; PubMed=2332426;  
 RA Moo H.-J., Shaw L.M., Messier J.M., Mercurio A.M.;  
 RT "The major non-integrin laminin binding protein of macrophages is identical to carbohydrate binding protein 35 (Mac-2)."; J. Biol. Chem. 265:7097-7099(1990).  
 RN [5]  
 RP DISULFIDE BOND.  
 RC MEDLINE=92011585; PubMed=1917966;  
 RA Moo H.-J., Lotz M.M., Tung J.U., Mercurio A.M.;  
 RT "Carbohydrate-binding protein 35 (Mac-2), a laminin-binding lectin, forms functional dimers using cysteine 186."; J. Biol. Chem. 266:18419-18422(1991).  
 CC -1- FUNCTION: Galactose-specific lectin which binds IGF.  
 CC -1- SUBUNIT: Exists as several distinct species of 35, 67 and 80 kDa. Probably forms homo- or heterodimers.  
 CC -1- TISSUE SPECIFICITY: The highest levels are found in activated macrophages.  
 CC -1- SIMILARITY: In the C-terminal section, belongs to the galactin (galactin/3-lectin) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X16834; CAA34736.1; -  
 CC EMBL: J03723; AAA37311.1; -  
 CC EMBL: X16074; CAA34206.1; -  
 CC PIR: S08537; A28651.  
 CC HSSP: P17931; 1A3K.  
 CC MGD: MGI:96778; Lgal3.  
 CC GO: GO:0005737; Cytoplasm; IDA.  
 CC GO: GO:0005515; F:protein binding; IPT.

DR InterPro: IPR008985; ConA\_like\_lect\_gl.  
 DR InterPro: IPR01079; Galectin.  
 DR Pfam: PF00337; Gal\_bind\_lectin; 1.  
 DR SMART: SM00276; GLECT; 1.  
 DR PROSITE: PS00309; GALAPTIN; 1.  
 KW Galectin; Lectin; IGE-binding protein; Repeat; Phosphorylation; Acetylation.  
 FT INIT MET 0 0  
 FT MOD RES 1 1  
 FT MOD RES 5 5  
 FT FT  
 FT DOMAIN 34 113  
 FT FT  
 FT REPEAT 34 42  
 FT REPEAT 43 51  
 FT REPEAT 52 60  
 FT REPEAT 61 69  
 FT REPEAT 70 78  
 FT REPEAT 79 87  
 FT REPEAT 88 96  
 FT REPEAT 97 106  
 FT REPEAT 107 113  
 FT DOMAIN 131 263  
 FT DISULFID 186 186  
 FT BINDING 194 200  
 FT CONFLICT 1 1  
 FT CONFLICT 3 3  
 FT CONFLICT 91 92  
 FT CONFLICT 109 111  
 FT CONFLICT 251 251  
 SQ SEQUENCE 263 AA; 27384 MW; 76C78AA0810D6EE CRC64;  
 Query Match 79.4%; Score 1077; DB 1; Length 263;  
 Best Local Similarity 77.9%; Pred. No. 3e-60;  
 Matches 208; Conservative 16; Mismatches 17; Indels 26; Gaps 8;  
 QY 2 ADNFSLDHALSSGSGNPNPGWPGAMGNGPAGAGYVPGAYPG-YPGAPPGAYPGQAPPG 60  
 DB 1 ADFSLNDHALSSGSGNPNPGYVPGAMGNGP-GAGGYVPGAYVPGAPPGAYPGQAPPG 59  
 QY 61 AYHG-----APGAYPGAPAPGVVPPSPGSGAYSSGQSPAPGAY-----A 101  
 DB 60 AYPGQAPPSAPVPGPTVPGAYVPGTAPGAYVPGQPA-PGAPP-GQPGAPGVPPCGSGAYPA 116  
 QY 102 TEPYGPAPPLVPPVPLPFGGVPRMLTTIGTVKPNANRALDPQRGDVAFFR-PFR 160  
 DB 117 AGPYGPAPPLVPPVPLPFGGVPRLTTIGTVKPNANRVLDPFRGDVAFFHPPFR 176  
 QY 161 NNNRRVIVCNKTLDDNNMGRSEEROSVFPESGKPFKIQVLEPDHFKVAVNDAAH-LQYXH 219  
 DB 177 NNNRRVIVCNKTKDDNNMGRSEEROSVFPESGKPFKIQVLEPDHFKVAVNDAAHLLQYXH 236  
 QY 220 RVKLTNEISKIGISGDIDLTLSASYTWI 246  
 DB 237 RKNKLRHSIQLSIGSDITLTLSANHAMI 263  
 RESULT 4  
 LEE3 RAT STANDARD; PRT; 261 AA.  
 AC P08699;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Galactin-3 (Galactose-specific lectin 3) (MAC-2 antigen) (IGE-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (CBP 35) (Laminin-binding protein) (Lectin L-29).  
 DE IGAUS3.  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxId=10116;  
 RN [1]



RP SEQUENCE FROM N.A.  
 RX MEDLINE=88016189; PubMed=2958848;  
 RA Albrandt K., Orida N.K., Liu F.-T.;  
 RT "An Igb-binding protein with a distinctive repetitive sequence and  
 RL homology with an Igc receptor."; Proc. Natl. Acad. Sci. U.S.A. 84:6859-6863(1987).  
 RN [2]  
 RP SEQUENCE OF 124-261 FROM N.A.  
 RX MEDLINE=85216641; PubMed=3858867;  
 RA Liu F.-T., Albrandt K., Mendel E., Kulczycki A. Jr., Orida N.K.;  
 RT "Identification of an Igc-binding protein by molecular cloning."; Proc. Natl. Acad. Sci. U.S.A. 82:4100-4104(1985).  
 RN [3]  
 RP SEQUENCE OF 119-144.  
 RX MEDLINE=90105471; PubMed=2605254;  
 RA Jelefer H., Masiarz P.R., Barondes S.H.;  
 RT "Soluble lactose-binding vertebrate lectins: a growing family."; Biochemistry 28:9222-9229(1989).  
 RN [4]  
 RP PARTIAL SEQUENCE, AND ACETYLATION.  
 RX MEDLINE=94075368; PubMed=8253805;  
 RA Hermann J., Turck C.W., Archison R.E., Huflejt M.E., Poulter L., Gilt M.A., Burlingame A.L., Barondes S.H., Jelfler H.;  
 RT "Primary structure of the soluble lactose binding lectin L-29 from rat and dog and interaction of its non-collagenous proline-, glycine-, tyrosine-rich sequence with bacterial and tissue collagenase."; J. Biol. Chem. 268:26704-26711(1993).  
 CC -1- FUNCTION: Galactose-specific lectin which binds IGE.  
 CC -1- SUBUNIT: Probably forms homo- or heterodimers.  
 CC -1- SIMILARITY: In the C-terminal section; belongs to the galectin (galaplin/S-lectin) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL, J02962; AAA40828.1; -;  
 DR EMBL, M13697; AAA41378.1; -;  
 DR PIR, A54889; A54889.  
 DR HSSP, P17931; 1A3X.  
 DR InterPro, IPR008985; Cona\_1like\_lect\_gl.  
 DR InterPro, IPR001079; Galectin.  
 DR Pfam, PF00337; Gal-bind\_lectin; 1.  
 DR SMART, SM00276; GLECT; 1.  
 DR PROSITE, PS00309; GALAPTIN; 1.  
 KM Galectin; lectin; Igc-binding protein; Repeat; Phosphorylation;  
 KW Acetylation.  
 FT INIT\_MET 0 0  
 FT MOD\_RES 1 1  
 FT MOD\_RES 5 5  
 FT DOMAIN 34 111  
 FT REPEAT 34 42  
 FT REPEAT 43 51  
 FT REPEAT 52 60  
 FT REPEAT 61 69  
 FT REPEAT 70 78  
 FT REPEAT 79 87  
 FT REPEAT 88 97  
 FT REPEAT 98 104  
 FT REPEAT 105 111  
 FT DOMAIN 129 261  
 FT DISULFID 184 184  
 FT BINDING 192 198  
 FT CONFLICT 19 19  
 SQ SEQUENCE 261 AA, 27070 KW, EAPFALVFSEA5080D CRC64;

Query Match 79.1%; Score 1074; DB 1; Length 261;  
 Best Local Similarity 77.9%; Pred. No. 4.6e-60;  
 Matches 204; Conservative 16; Mismatches 24; Indels 18; Gaps 7;  
 QY 2 ADNSLHDALSGSNPNQCPGAWGNOPAGAGYPCASITG-YPGCAPPAAYRQAPRG 60  
 DB 1 ADGFSLNDALAGSNPNQCPGAWGNQP-GAGGYPCASYGAYPGQAPPGGYRQAPPS 59  
 QY 61 AYHG-----APGAYGAPAGYVPGPGPGGAYPSS--GQSPAGAY-ATGPGY 106  
 DB 60 AYPGPTGSAVPGPTAPGAPGPTAPGAFPGQPGGPAVPSAPGAYPAGYATGTFPG 119  
 QY 107 APAGELIVPNLPLPGGVVPRMLITLIGTVKPNNRIALDFQGNDAVHFH-PRNENNR 165  
 DB 120 APTGELIVPNMPLPGGVMPRLITLIGTVKPNNSITLNKKGDIAFHFNPRNENNR 179  
 QY 166 RVIYCNITKLDNNKGEERQSYPPESGCKPKIQVIVEDPHKVAVNDNH-IQYHRYVKL 224  
 DB 180 RVIYCNTRQDNNMGGEERQSAFPESGCKPKIQVIVEDHFKVAVNDVHLIQYHNRKNTL 239  
 QY 225 NEISKLTGSSPIDLTSASTMI 246  
 DB 240 REISQLGITGITLTLSASHMI 261  
 RESULT 5  
 LEG3 CRITO STANDARD; PRT; 244 AA.  
 AC P47953;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Galectin-3 (Galactose-specific lectin 3) (MNC-2 antigen) (Igc-binding protein) (35 kDa lectin) (carbohydrate binding protein 35) (CBP 35) (Laminin-binding protein) (Lectin L-29) (CBP30).  
 GN LGALS3.  
 OS Cricetulus longicaudatus (long-tailed hamster) (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 OX NCBI\_TaxID=10030;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94299546; PubMed=8027086;  
 RA Mehrl B., Bawumia S., Martin S.R., Hughes R.C.;  
 RT "Structure of baby hamster kidney carbohydrate-binding protein CBP30, an S-type animal lectin."; J. Biol. Chem. 269:18250-18258(1994).  
 RL J. Biol. Chem. 269:18250-18258(1994).  
 CC -1- FUNCTION: Galactose-specific lectin which binds IGE.  
 CC -1- SIMILARITY: In the C-terminal section; belongs to the galectin (galaplin/S-lectin) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL, X78879; CAA55479.1; -;  
 DR HSSP, P17931; 1A3X.  
 DR InterPro, IPR008985; Cona\_1like\_lect\_gl.  
 DR InterPro, IPR001079; Galectin.  
 DR Pfam, PF00337; Gal-bind\_lectin; 1.  
 DR SMART, SM00276; GLECT; 1.  
 DR PROSITE, PS00309; GALAPTIN; 1.  
 KM Galectin; lectin; Igc-binding protein; Repeat; Phosphorylation;  
 KW Acetylation.  
 FT INIT\_MET 0 0  
 FT MOD\_RES 1 1  
 FT MOD\_RES 5 5  
 FT SEQUENCE 261 AA, 27070 KW, EAPFALVFSEA5080D CRC64;

```

      (BY SIMILARITY).
      7 X 9 AA TANDEM REPEATS OF Y-P-G-X(3) -P-
      [GS] -A.
      FT DOMAIN 34 98
      FT REPEAT 34 42
      FT REPEAT 43 51
      FT REPEAT 52 60
      FT REPEAT 61 69
      FT REPEAT 70 77
      FT REPEAT 78 87
      FT REPEAT 88 98
      FT DOMAIN 112 244
      FT DISULFID 167 167
      FT BINDING 175 181
      SQ SEQUENCE 244 AA: 25608 MW, 8F99B9AA0BBA7D3F CRC64;

Query Match 77.0%; Score 1044.5; DB 1; Length 244;
Best Local Similarity 80.9%; Pred. No. 2.9e-58;
Matches 203; Conservative 12; Mismatches 23; Indels 13; Gaps 8;

QY 2 ADNFSLHDALSSGNPNPGMPCGAMGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPPG 60
DB 1 ADGFSINDALSSGNPNPGMPCGAMGNOP-GAGGTFGASYPGAYPGQAPPGAYPGQAPPG 59
QY 61 AYHG--APGAYPGAPAGYVPGPPSGGAYPSSGQSPAGAY-ATGPGAPAGPLIVPN 117
DB 60 AYGPPTAPGAYPG-PAPGAYPGQPGASGAY-----PSAPGAYPAGPYGAPGALTVPYK 113
QY 118 LPLPGSVPRMLITIGTYKENANRALDPQGNDAVHFH-PRFENNRRVIVCNTKLDN 176
DB 114 LPLAGVPMRMLITIGTYKENANRIILNFRGNDAVHFHPRFENNRRVIVCNTKLDN 173
QY 177 NMGREKQSVPPESGPKFQIVLVEPDHFKVAVMDAH-LQYNHRVKKLNIISKLIGSD 235
DB 174 NMGREKQAFPPESGPKFQIVLVEADHFKVAVMDAHLLQYNHRKMLREINQMEISGD 233
QY 236 IDLTASATYMT 246
DB 234 ILLTSAATMT 244

RESULT 6
LEG3 RABIT
ID LEG3 RABIT STANDARD; PRT; 241 AA.
AC P47845;
DB 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-3 (Galactose-specific lectin 3) (Mac-2 antigen) (Igf-binding protein) (35 kDa lectin) (Carbohydrate binding protein 35) (CBP 35)
DE (Laminin-binding protein) (Lectin L-29).
DE IGALS3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Aorta;
RX MEDLINE=96011642; PubMed=7590275;
RA Gaudin J.-C., Monsigny M., Legrand A.,
RT "Cloning of the cDNA encoding rabbit galectin-3."
RL Gene 163:249-252(1995).
CC -1- FUNCTION: Galactose-specific lectin which binds IgE.
CC -1- SIMILARITY: In the C-terminal section, belongs to the galectin (galactin/5-lectin) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

```

```

      EMBL; U06470; AAC8491.1; -.
      DR F1R; J04300; J04300.
      DR HSSP; P17931; IAKK.
      DR InterPro; IPR008985; Cona1_1ec_g1.
      DR InterPro; IPR001079; Galectin.
      DR Pfam; PF00337; Gal-bind_lectin; 1.
      DR SMART; SM00276; GLECT_1.
      DR PROSITE; PS00309; GALAPTIN; 1.
      DR Galectin; Lectin; Igf-binding protein; Repeat; Phosphorylation;
      KW Acetylation.
      FT INIT MET 0 0
      FT MOD_RES 1 1
      FT MOD_RES 5 5
      FT MOD_RES 5 5
      FT DOMAIN 34 97
      FT REPEAT 34 42
      FT REPEAT 43 51
      FT REPEAT 52 60
      FT REPEAT 61 69
      FT REPEAT 70 79
      FT REPEAT 80 91
      FT REPEAT 92 97
      FT DOMAIN 109 241
      FT DISULFID 164 164
      FT BINDING 172 180
      SQ SEQUENCE 241 AA: 25370 MW, 6EFC42D626479F9 CRC64;

Query Match 75.1%; Score 1019; DB 1; Length 241;
Best Local Similarity 78.7%; Pred. No. 1.1e-56;
Matches 196; Conservative 15; Mismatches 26; Indels 12; Gaps 7;

QY 2 ADNFSLHDALSSGNPNPGMPCGAMGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPPG 60
DB 1 ADGFSINDALSSGNPNPGMPCGAMGNOPAGAGYPGAYPGAYPGAH-PGAYPGQAPPG 59
QY 61 AYHGAPGAYPGAPAGYVPGPPSGGAYPSSGQSPAGAY-ATGPGAPAGPLIVPN 119
DB 60 PYPG-PGAA-----GAYPGQPGAYPSPGQSPAGAYPGASPSASAGPLVYYDLP 112
QY 120 LPLGVPRLMLITIGTYKENANRIALDPQGNDAVHFH-PRFENNRRVIVCNTKLDN 178
DB 113 LPLGVPRLMLITIGTYKENANRIALDPQGNDAVHFHPRFENNRRVIVCNTKLDN 172
QY 179 GREEKQSVPPESGPKFQIVLVEPDHFKVAVMDAH-LQYNHRVKKLNIISKLIGSD 237
DB 173 GREEKQTFPPFQIGKPKFQIVLVEPDHFKVAVMDAHLLQYNHRMRLKEINKLIGSDIQ 232
QY 238 LTSASATYMT 246
DB 233 LTSASATMT 241

RESULT 7
LEG3 MOUSE
ID LEG3 MOUSE STANDARD; PRT; 353 AA.
AC C08573; C08572;
DB 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-9.
DE IGALS9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney, and Small intestine;
RX MEDLINE=97190351; PubMed=9038233;
RA Wada Y., Kanwar Y.S.;
RT "Identification and characterization of galectin-9, a novel beta-

```

```

RT      galactoside-binding mammalian lectin."
RL      J. Biol. Chem. 272:6078-6086(1997).
RX      CHARACTERIZATION.
RX      MEDLINE=97298141; PubMed=9153289;
RA      Wada T., Ota K., Kumar A., Wallner E.I., Kanwar Y.S.;
RT      "Developmental regulation, expression, and apoptotic potential of
RT      galectin-9, a beta-galactoside binding lectin."
RL      J. Clin. Invest. 99:2452-2461(1997).
CC      -1- FUNCTION: Binds galactosides. May play a role in thymocyte-
CC      epithelial interactions relevant to the biology of the thymus.
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-
CC      CLASSICAL SECRETORY PATHWAY.
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Comment=Additional isoforms seem to exist;
CC      Name=Long;
CC      IsoId=O08573-1; Sequence=Displayed;
CC      Name=Short;
CC      IsoId=O08573-2; Sequence=VSP_003097;
CC      -1- TISSUE SPECIFICITY: Accentuated expression in liver and thymus of
CC      embryo, detected in embryonic heart, brain, lung, liver, and
CC      kidney. Highly expressed in adult thymus, small intestine, and
CC      liver, and to a lesser extent in lung, kidney, spleen, cardiac,
CC      and skeletal muscle. Barely detectable in brain and reticulocyte.
CC      The long form is expressed exclusively in the small intestine.
CC      -1- DEVELOPMENTAL STAGE: The expression increased with successive
CC      stages of embryonic development.
CC      -1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
CC      domains.
CC      -1- SIMILARITY: Belongs to the galectin (galaptin/s-lectin) family.
CC      CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcements
CC      or send an email to license@isb-sib.ch).
CC      CC
CC      EMBL; U55061; AAB51189.1; -
CC      EMBL; U55060; AAB51189.1; -
CC      HSSP; P17931; 1A3K.
CC      DR      WGD; WGI:109496; LGal9s.
CC      DR      InterPro; IPR008985; Consa like_lect_g1.
CC      DR      InterPro; IPR001079; Galectin.
CC      DR      Pfam; PF00337; Gal-bind_lectin; 2.
CC      DR      SMART; SM00276; GLECT; 2.
CC      DR      PROSITE; PS00309; GALAPTIN; 2.
CC      KM      Galectin, Lectin, Repeat; Alternative splicing.
CC      FT      DOMAIN 1 147
CC      FT      DOMAIN 148 204
CC      FT      BINDING 205 353
CC      FT      BINDING 81 87
CC      FT      BINDING 285 291
CC      FT      VAREPLIC 148 178
CC      FT      /FTId=VSE_003057.
CC      SQ      SEQUENCE 353 AA; 40036 MW; B54036F6E280C531 CRC64;
CC
CC      Query Match 32.0%; Score 299; DB 1; Length 353;
CC      Best Local Similarity 37.5%; Pred. No. 4,7e-12;
CC      Matches 69; Conservative 28; Mismatches 47; Indels 40; Gaps 8
CC
CC      63 HGAGC--AYGAPAPGVYPPGPPSGGAVYSSGQSPAPGAYATGPGAGGLIYVYNP 119
CC      DB HSTPQGMSTPEIP-PVYVFP-----AYT-----IPFYTP 228
CC
CC      120 LAGGVAPRMILITLGVKKNAMRIALDPORGNDVAFH--PRENNNNRYVYNTKLDNNW 178
CC      DB IPNGLYPSSIMTSGVLPDARFRHINLRCGGDIAPHLNPRNEN--AVANTQTINSM 285
CC
CC      179 GEEHQSV--PPFESGKPEKIVLEPDHFKVAVNDAMH-QYNHRVKKLNEISKLGISGD 235

```

```

DB      286  GQERSLTGMPEFSGGSFWIITCEBCHKVAVNGHMCXEYHRLKNLDINTLVEAGD 345
QY      236  IDLT 239
        |||
Db      346  IOLT 349

RESULT 8
LEG9_HUMAN STANDARD; PRT; 355 AA.
ID      LEG9_HUMAN STANDARD; PRT; 355 AA.
AC      000182; 014532; 075028; GONG08;
DT      01-NOV-1997 (Rel. 35, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Galectin-9 (HOM-HD-21) (Ecalectin).
GN      LGALS9.
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NC      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM SHORT).
RC      TISSUE=Spleen;
RX      MEDLINE=97197815; PubMed=9045665;
RA      Tuerect O., Schmitt H., Fadle N., Pfreundschuh M., Sahin U.;
RT      "Molecular definition of a novel human galectin which is immunogenic
RT      in patients with Hodgkin's disease.";
RL      J. Biol. Chem. 272:6416-6422(1997).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORM LONG).
RC      TISSUE=Gastric carcinoma;
RA      Kato S.;
RL      Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A. (ISOFORM SHORT).
RC      MEDLINE=98307937; PubMed=9642261.
RA      Matsunoto R., Masunoto H., Seki M., Hata M., Asano Y., Kanegasaki S.,
RA      Stevens R.L., Hiraishima M.;
RT      "Human ecalectin, a variant of human galectin-9, is a novel eosinophil
RT      chemoattractant produced by T lymphocytes.";
RL      J. Biol. Chem. 273:16976-16984(1998).
RN      [4]
RP      SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RA      Akiyama S.;
RT      "Homo sapiens galectin-9 (LGALS9) / ecalectin gene, exon 2 through
RT      11.";
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN      [5]
RP      SEQUENCE FROM N.A. (ISOFORM LONG).
RA      Graessler J., Spitzemberger F., Schroeder H.E.;
RT      "Genomic organization of the human galectin-9 gene.";
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: Binds galactosides. May play a role in thymocyte-
CC      epithelial interactions relevant to the biology of the thymus. The
CC      short isoform acts as an eosinophil chemoattractant.
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Comment=Additional isoforms seem to exist;
CC      Name=Long;
CC      IsoId=000182-1; Sequence=Displayed;
CC      Name=Short;
CC      IsoId=000182-2; Sequence=VSP 003096;
CC      -1- TISSUE SPECIFICITY: Peripheral blood leukocytes and lymphatic
CC      tissues. Overexpressed in Hodgkin's disease tissue.
CC      -1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
CC      domains.
CC      -1- SIMILARITY: Belongs to the galectin (galactin/s-lec) family.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial

```

RESULT 9		
LEG9_RAT	STANDARD;	PRT; 354 AA
ID LEG9_RAT		
AC P97840; 008588; 035866;		
DT 01-NOV-1997 (Rel. 35, Created)		

DT	15-JUN-1998 (Rel. 35, last sequence update)
DT	15-MAR-2004 (Rel. 43, last annotation update)
DE	Galectin-9 (36 kDa beta-galactoside binding lectin) (Urate transporter/channel) (UAT).
DE	lgALS9.
GN	lgALS9.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eubacteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC	STRAIN=Sprague-Dawley; TISSUE=Intestine;
RX	MEDLINE=97190351; PubMed=9038233;
RA	Meda J., Kanwar Y.S.;
RA	Leal-Pinto E., Tao W., Rappaport J., Richardson M., Knorr B.A.,
RA	Abrahamson R.G.;
RT	"Molecular cloning and functional reconstitution of a urate transporter/channel.";
RT	J. Biol. Chem. 272:617-625 (1997).
CC	-1- FUNCTION: Binds galactosides. May play a role in thymocyte-epithelial interactions relevant to the biology of the thymus (By similarity). May provide the molecular basis for urate flux across cell membranes, allowing urate that is formed during purine metabolism to efflux from cells and serving as an electrogenic transporter that plays an important role in renal and gastrointestinal urate excretion. Highly selective to the anion urate.
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-CLASSICAL SECRETORY PATHWAY (By similarity).
CC	-1- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Comment=Additional isoforms seem to exist;
CC	Name=Long;
CC	Isoid=P97840-1; Sequence=Displayed;
CC	Name=Short;
CC	Isoid=P97840-2; Sequence=VSP_003098;
CC	-1- TISSUE SPECIFICITY: The long form is expressed exclusively in the small intestine.
CC	-1- DOMAIN: Contains two homologous but distinct carbohydrate-binding domains.
CC	-1- SIMILARITY: Belongs to the galectin (galactin/S-lectin) family.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
CC	EMBL; U59462; AAB51192.1; -;
CC	EMBL; U72741; AAB68592.1; -;
DR	EMBL; U67958; AAB48591.1; -;
DR	HSSP; P17931; 1A3K.
DR	InterPro; IPR008985; Cona_like_lect_51.
DR	InterPro; IPR01079; Galectin.
DR	Pfam; PF00337; Gal-bind_lectin; 2.
DR	SMART; SM00276; GLECY; 2.
DR	PROSITE; PS00309; GALACTIN; 2.
KM	Galectin; Lectin; Repeat; Alternative splicing; Ion transport.
FT	DOMAIN 1 147
FT	DOMAIN 148 205 LINKER
FT	DOMAIN 206 354 GALACTIN 2.
FT	BRINDING 81 87 BETA-GALACTOSIDE (By SIMILARITY).
FT	BRINDING 286 292 BETA-GALACTOSIDE (By SIMILARITY).
FT	BRINDING 148 179 Missing (in isoform short).
FT	VARSPLIC



```

DR HSSP; P47929; 1BKZ.
DR InterPro; IPR008985; Cons. like _lec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
DR Galectin; Lectin; Acetylation.
KW Galectin; Lectin; Acetylation.
FT INIT MET 0
FT MOD RES 1 1 ACETYLATION.
FT BINDING 76 82 BETA-GALACTOSIDE (POTENTIAL).
FT CONFLICT 127 135 MISSING (IN REF. 1; AA65445).
SQ SEQUENCE 144 AA; 16065 MW; BC95283D760DA515 CRC64;

Query Match 18.5%; Score 250.5; DB 1; Length 144;
Best Local Similarity 40.4%; Pred. No. 2.1e-09;
Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;

QY 108 PAGPLIVPNLPLPGGVPRMLTITIGTVKPNANRALDFOGNDVAEHP-PRFNNR 166
DB 8 PYVNLAVPFTSIPNGLYPSKSIIVSGVLSDAKRFQINRCGGDIAFHLNPRFEN-- 64

QY 167 VVYCNKLDNNMGERERQ--SVFPESGKRFKQVLPVPHFVAVNDAL-QYNNRVK 223
DB 65 AAVRNTQINNSWGPERRSLPGSNPFSRGGFVSVMILCEGHCFKAVADGCHI CEYSHLMN 124

QY 224 LNEISKIGISGDIIDL 239
DB 125 LPDINTLLEVAQDIFLT 140

RESULT 12
LEG6 MOUSE STANDARD; PRT; 301 AA.
ID LEG6_MOUSE 054851; O88352;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Galectin-6.
GN LGALS6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98112847; PubMed=9446608;
RA Gilt M.A., Colnot C., Potlter F., Nani K.J., Barondes S.H.,
RA Leflier H.;
RT "Galectin-4 and galectin-6 are two closely related lectins expressed
RT in mouse gastrointestinal tract.";
RL J. Biol. Chem. 273:2954-2960(1998).

RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98112848; PubMed=9446609;
RA Gilt M.A., Xia Y.-R., Atchison R.E., Lusis A.J., Barondes S.H.,
RA Leflier H.;
RT "Sequence, structure, and chromosomal mapping of the mouse Lgal6
RT gene, encoding galectin-6.";
RL J. Biol. Chem. 273:2961-2970(1998).

-1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
-1- domains.
-1- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; AF026799; AAC04508.1; -.
DR EMBL; AF026796; AAC04508.1; JOINED.
DR EMBL; AF026797; AAC04508.1; JOINED.
DR EMBL; AF026798; AAC04508.1; JOINED.
DR EMBL; AF026794; AAC27244.1; -.
DR HSSP; P17931; LA3K.
DR MGD; MGI:107535; Lgal6.
DR InterPro; IPR008985; Cons. like _lec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
DR Galectin; Lectin; Repeat.
KW Galectin; Lactose-binding protein.
FT DOMAIN 1 151 GALAPTIN 1.
FT DOMAIN 152 160 LINKER.
FT DOMAIN 160 301 GALAPTIN 2.
FT CONFLICT 154 154 A -> V (IN REF. 2; AAC27244).
SQ SEQUENCE 301 AA; 34112 MW; 9A4DD09944EDFAB9 CRC64;

Query Match 18.1%; Score 245; DB 1; Length 301;
Best Local Similarity 37.7%; Pred. No. 8.8e-09;
Matches 57; Conservative 29; Mismatches 55; Indels 10; Gaps 6;

QY 99 AVATGYGAPAGPLIVPNLPLPGGVPRMLTITIGTVKPNANRALDF---QGNDA 154
DB 2 AVVAPAGYQPTYNPLTPYKRPILPGSLVGMSPFYIGTAKENMRFFVNPVAVGDDADAVA 61

QY 155 FHP-PRFNNRNVYCNKLDNNMGERERQSVFPESGKRFKQVLPVPHFVAVNDVA 213
DB 62 FHPNPRFDGMDK--VVENTKQSGRMGKEBKSN-PRQKGHFLVPMVEHYKVVNNS 118

QY 214 -HLYNNHRYVKLNEISKIGISGDIIDLTSASY 243
DB 119 PEVEYHRL-PYQVMTHLQVDDGLDLSINF 148

RESULT 13
LEG4 RAT STANDARD; PRT; 324 AA.
ID LEG4_RAT 038552;
AC 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE (L36LBP).
GN LGALS4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=93194902; PubMed=8449956;
RA Oda Y., Herrmann U., Gilt M., Turck C.W., Burlingame A.L.,
RA Barondes S.H., Leflier H.;
RT "Soluble lactose-binding lectin from rat intestine with two different
RT carbohydrate-binding domains in the same peptide chain.";
RL J. Biol. Chem. 268:5929-5939(1993).

-1- FUNCTION: GALACTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
-1- SUGARS.
-1- SUBUNIT: Monomer.
-1- TISSUE SPECIFICITY: Highly expressed in full-length form in small
-1- and large intestine and stomach but was not detected in other
-1- tissues including lung, liver, kidney and spleen.
-1- DOMAIN: Contains two homologous but distinct carbohydrate-binding
-1- domains.
-1- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```





CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: I78132; AAB51605.1; -  
DR EMBL: X91790; CAA62904.1; ALT\_INIT.  
DR EMBL: AF074000; AAD45402.1; -  
DR EMBL: AF074001; AAD45403.1; -  
DR EMBL: AF074002; AAD45404.1; -  
DR EMBL: AF193806; AAF19370.1; ALT\_INIT.  
DR EMBL: AF193805; AAF19370.1; JOINED.  
DR EMBL: AF342815; AAK16735.1; ALT\_INIT.  
DR EMBL: AL136105; CAC15946.1; ALT\_INIT.  
DR EMBL: AL136105; CAC15947.1; ALT\_INIT.  
DR EMBL: BC018618; AAI18618.1; -  
DR EMBL: BC016486; AAI16486.1; ALT\_INIT.  
DR HSSP: P17931; IAK3K.  
DR Genew; HGNC:6569; LGALS8.  
DR MIM: 606099; -  
DR GO: GO:0005615; C:extracellular space; TAS.  
DR GO: GO:0005530; F:lectin; TAS.  
DR InterPro: IPR008985; Cona\_like\_lect\_gl.  
DR InterPro: IPR001079; Galectin.  
DR Pfam: PF00337; Gal-bind lectin; 2.  
DR PROSITE: PS00309; GALAPTIN; 1.  
KW Galectin; Lectin; Repeat; Antigen; Alternative splicing; Polymorphism.  
FT DOMAIN 1 153 GALAPTIN 1.  
FT DOMAIN 154 184 LINKER.  
FT DOMAIN 185 316 GALAPTIN 2.  
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).  
FT VARSPPLIC 182 182 L -> LPSNRGDISKIAPIYVTKSKDSVNHNTLCTKIP  
PMNVSK (in isoform 2).  
FT FT  
FT VARSPPLIC 182 182 /FTID=VSP\_003094.  
L -> LQVSPSPMDLQSGSEFTFCVLMTRVLEIAFCRPI  
GLTVASRQ (in isoform 3).  
FT FT  
FT VARIANT 18 18 /FTID=VSP\_003095.  
Y -> F.  
FT FT  
FT VARIANT 35 35 C -> R.  
FT FT  
FT VARIANT 55 55 /FTID=VAR\_009710.  
V -> M.  
FT FT  
FT CONFLICT 97 99 /FTID=VAR\_012991.  
KRE -> OKER (IN REF. 2).  
FT FT  
FT CONFLICT 111 111 D -> A (IN REF. 2).  
FT FT  
FT CONFLICT 170 170 S -> V (IN REF. 1).  
FT FT  
FT CONFLICT 183 183 S -> R (IN REF. 2, 6, 7 AND 8).  
FT FT  
FT CONFLICT 203 203 K -> Q (IN REF. 1).  
SQ SEQUENCE 316 AA; 35538 MW; 1724C36085A0510A CRC64;  
Query Match 16.8%; Score 228.5; DB 1; Length 316;  
Best Local Similarity 40.0%; Pred. No. 9.7e-08;  
Matches 56; Conservative 22; Mismatches 51; Indels 11; Gaps 5;  
QY 113 IPEYNPLPFGGVPRMLITLIGTKPNANRIADFGNGN-----DVAHFH-PRFENNRR 165  
DB 15 VLPYVGTITPDQDPLGLIVICGHVPSDADRFQVDLQNGSSVKPRAVAFHFNRRFRAG- 73  
QY 166 RYIVCNCTKLDNNMGREERQSVFPFESGPKPIQVLVEPPEFKYAVNDAA-LQYNERVKUL 224  
DB 74 -CIVCNLTLLNEKKGMEIITDTPPKREKSEFVIYMLKQFQVAVNGKHTLLYGRIGP- 131  
QY 225 NEISKLGISGIDLTLSASYT 244  
DB 132 EKIDTIGIFGKVNINHSIGF 151

RESULT 15  
LEG8 RAT STANDARD; PRT; 316 AA.  
AC 062665;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Galectin-8 (30 kDa S-type lectin) (RL-30).  
GN LGALS8.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=liver;  
RX MEDLINE=95155445; PubMed=7852431;  
RA Hadadit Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y.;  
RT Galectin-8, A new rat lectin, related to galectin-4.";  
RL J. Biol. Chem. 270:3447-3453(1995).  
CC - FUNCTION: Possesses sugar binding and hemagglutination activity.  
CC - SUBCELLULAR LOCATION: Cytoplasmic.  
CC - TISSUE SPECIFICITY: Expressed in liver, kidney, cardiac muscle,  
CC lung, and brain.  
CC - DEVELOPMENTAL STAGE: Very low levels in whole embryos, high levels  
CC in adult tissues.  
CC - DOMAIN: Contains two homologous but distinct carbohydrate-binding  
CC domains.  
CC - SIMILARITY: Belongs to the galectin (galaptin/s-lectin) family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL, contribution  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U09824; AAB6359.1; -  
DR PIR: A55975; A55975.  
DR HSSP: P17931; IAK3K.  
DR InterPro: IPR008985; Cona\_like\_lect\_gl.  
DR InterPro: IPR001079; Galectin.  
DR Pfam: PF00337; Gal-bind lectin; 2.  
DR SMART: SM00276; GLECF; 2.  
DR PROSITE: PS00309; GALAPTIN; 1.  
KW Galectin; Lectin; Repeat.  
FT DOMAIN 1 153 GALAPTIN 1.  
FT DOMAIN 154 184 LINKER.  
FT DOMAIN 185 316 GALAPTIN 2.  
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).  
SQ SEQUENCE 316 AA; 36038 MW; C04B76C6FB913D59 CRC64;  
Query Match 16.1%; Score 218.5; DB 1; Length 316;  
Best Local Similarity 39.9%; Pred. No. 4e-07;  
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;  
QY 114 VEPNPLPFGGVPRMLITLIGTKPNANRIADFGNGN-----DVAHFH-PRFENNRR 166  
DB 16 IPEVSTITBQLRGSLIVIRGHVPEKDSERFQVDFQGNLKPADVAFFHFNRRFKSN-- 73  
QY 167 VYIVCNCTKLDNNMGREERQSVFPFESGPKPIQVLVEPPEFKYAVNDAA-LQYNERVKUL 225  
DB 74 CIVCNLTLLNEKKGMEIITDTPPKREKSEFVIYMLKQFQVAVNGKHTLLYAHRLNP-E 132  
QY 226 EISKLGISGIDLTLSASY 243  
DB 133 KIDTIGIFGKVNINHSIGF 150

RESULT 16  
LEG4 HUMAN STANDARD; PRT; 323 AA.  
AC P56470;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)  
DE (L36LBP) (Antigen NY-CO-27).





RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hutylk S.W.,  
 RA Villalón D.K., Mundy D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahney U., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Maria M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RA and mouse cDNA sequences.";  
 RA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RL [4]  
 RP FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=21671326; PubMed=11706006;  
 RA Kuwabara I., Kuwabara Y., Yang R.Y., Schuler M., Green D.R.,  
 RA Zuraw B.L., Han D.K., Liu F.T.;  
 RT "Galectin-7 (PGI) exhibits pro-apoptotic function through JNK  
 RT activation and mitochondrial cytochrome c release.";  
 RL J. Biol. Chem. 277:3487-3497 (2002).  
 RL [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=98434364; PubMed=9760227;  
 RA Leonidas D.D., Vatzaki E.H., Vorun H., Celis J.E., Madsen P.,  
 RA Acharya K.R.;  
 RT "Structural basis for the recognition of carbohydrates by human  
 RT galectin-7.";  
 RL Biochemistry 37:13930-13940 (1998).  
 CC -I- FUNCTION: Could be involved in cell-cell and/or cell-matrix  
 CC interactions necessary for normal growth control. Pro-apoptotic  
 CC protein that functions intracellularly upstream of JNK activation  
 CC and cytochrome c release.  
 CC -I- SUBUNIT: Monomer.  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic and nuclear. May be secreted by  
 CC a non-classical secretory pathway.  
 CC -I- TISSUE SPECIFICITY: Mainly in stratified squamous epithelium.  
 CC -I- INDUCTION: By p53.  
 CC -I- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L07769; AAA67899.1; -;  
 DR EMBL; U06643; AAA68820.1; -;  
 DR EMBL; BC042911; AAH42911.1; -;  
 DR PIR; I55469; I55469.  
 DR PDB; 1BKZ; 04-NOV-98.  
 DR PDB; 2GAL; 04-NOV-98.  
 DR PDB; 3GAL; 04-NOV-98.  
 DR PDB; 4GAL; 04-NOV-98.  
 DR PDB; 5GAL; 04-NOV-98.  
 DR Aarhus/Chem-ZDPAGE; 17; IEF.  
 DR Genew; HGNC:6568; LGAL57.  
 DR MIM; 600615; -;  
 DR GO; GO:0005615; C:extracellular space; TAS.  
 DR GO; GO:0005530; F:lectin; TAS.  
 DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.  
 DR InterPro; IPR008985; Coma\_like\_lect\_1.  
 DR InterPro; IPR001079; Galectin.  
 DR Pfam; PF00337; Gal-bind\_lectin; 1.  
 DR SMART; SM00276; GLECF; 1.  
 DR PROSITE; PS00109; GALAPTIN; 1.  
 KM Galectin; Lactin; Apoptosis; Nuclear protein; 3D-structure.  
 FT INIT MET 0 0  
 FT BINDING 69 75 BETA-GALACTOSIDE (POTENTIAL).  
 FT STRAND 5 8

FT TURN 10 11  
 FT TURN 15 16  
 FT STRAND 18 25  
 FT TURN 27 28  
 FT STRAND 31 37  
 FT TURN 42 43  
 FT STRAND 46 53  
 FT STRAND 54 57  
 FT STRAND 58 65  
 FT TURN 66 67  
 FT STRAND 68 69  
 FT STRAND 73 74  
 FT TURN 82 83  
 FT STRAND 84 92  
 FT STRAND 96 101  
 FT TURN 102 103  
 FT STRAND 104 110  
 FT HELIX 115 117  
 FT STRAND 120 125  
 FT STRAND 128 135  
 SQ SEQUENCE 135 AA; 14944 MW; 4E7CEA54036EF806 CRC64;  
 Query Match 15.5%; Score 211; DB 1; Length 135;  
 Best Local Similarity 35.3%; Pred. No. 5.3e-07;  
 Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;  
 QY 114 VRYNPLPGVVPRLITLITGVKRNARIALDF----QRGNDVAFHF--PRENNRRVY 168  
 DB 3 VHKSSLPEGRIPGVRLRGVLPNNASRFVNLICGEGSGDALHFRPLDTS--V 59  
 QY 169 VCNITKLDNNKNGEEROSQVPEFSGKPKIQVLVEPDHFKVAVNDA-HIQYNRVKLANEI 227  
 DB 60 VNSKEGSGWGEERKPGVPGPGPFVLLIASDDGFAVVGDAQYHHFRRL-PLAVY 118  
 QY 228 SKLGISGIDILTS 240  
 DB 119 RLVEVGVDVLDLS 131  
 RESULT 18  
 LE68 MOUSE STANDARD; PRT; 316 AA.  
 AC 09Jul15;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Galectin-8 (LGALS-8).  
 GN LGALS8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RA Maier C., Haussler J., Roesch K., Moschgath E., Haussler J.,  
 RA Vogel W.;  
 RT "The human LGALS-8 gene: genomic sequence and expression of the  
 RT prostate carcinoma tumour antigen (PCTA-1) and the Pöse carbohydrate  
 RT binding protein.";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo W.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Possesses sugar binding and hemagglutination activity.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF218069; AAF27645.1; -;  
 DR EMBL: BC040243; AAH40243.1; -;  
 DR HSSP: P17931; IAXK.  
 DR MGD: MGI:1928481; Tgals8.  
 DR InterPro: IPR008985; Consa\_1like\_1ec\_g1.  
 DR InterPro: IPR01079; Galectin.  
 DR Pfam: PF00337; Gal-bind\_lectin; 2.  
 DR SMART: SM00276; GLECT; 2.  
 DR PROSITE: PS00309; GALACTIN; 1.  
 DR KW: Galectin; Lectin; Repeat.  
 FT DOMAIN 1 153 GALACTIN 1.  
 FT DOMAIN 154 184 LINKER.  
 FT DOMAIN 185 316 GALACTIN 2.  
 FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).  
 SQ SEQUENCE 316 AA; 36161 MW; 11A20309A5F52C69 CRC64;  
 Query Match 15.3%; Score 207.5; DB 1; Length 316;  
 Best Local Similarity 38.8%; Pred. No. 1.9e-06;  
 Matches 54; Conservative 19; Mismatches 55; Indels 11; Gaps 5;  
 QY 113 TVPNVLPDGGVPMPLITITIGVKNANRIALDFQGN-----DVAHF-PRPENNRR 165  
 DB 15 IIPVGTTEQLKPSLILVIRGHVPCSERFQVQLNSLKPRADVAFHENPREKSS- 73  
 QY 166 RVIVCNTLDNNWGEERQSVPEPSGPEKIQVYVEDHPRVAVNDAH-LQYNRRVKKL 224  
 DB 74 -CIVCNLTLOEKMGWEELTYDMPFKKESFELVFWLKNKQVAVNGRHVLLYARISP- 131  
 QY 225 NEISKLGISGIDILTSASY 243  
 DB 132 EQIDTVGIYKVNHSIGF 150  
 RESULT 19  
 CM32 HUMAN STANDARD; PRT; 245 AA.  
 AC Q8NDG0; Q96BG5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protein c14orf32.  
 GN C14ORF32.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12508121;  
 RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,  
 RA Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,  
 RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,  
 RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cnuad C.,  
 RA Bruns T., Jallion O., Friedlander L., Samson G., Brothier P.,  
 RA Cure S., Segreus B., Aniere F., Samin S., Crespeau H., Abbasi N.,  
 RA Alich N., Boscus D., Dickhoff R., Dots M., Dubois I., Friedman C.,  
 RA Gouyvenoux M., James R., Madan A., Maizy-Betarda B., Mangent S.,  
 RA Martins N., Menard M., Ozias S., Ratcliffe A., Shaffer T., Trask B.,  
 RA Vacherie B., Bellemere C., Belser C., Besnard-Gonnet M.,  
 RA Barcol-Mayel D., Boulard M., Brier-Silla S., Combette S.,  
 RA Dufosse-Laurent V., Perron C., Lechaplais C., Louesse C., Muselet D.,  
 RA Magdelanet G., Patieu E., Petit R., Sirvain-Trukniewicz P., Trypou A.,  
 RA Vega-Carny N., Bataille E., Bluet E., Borel-Lais I., Dubois M.,  
 RA Dumont C., Guerin T., Haffray S., Hammadi R., Muanga J., Pellouin V.,  
 RA Robert D., Wunderle E., Gauguier G., Roy A., Sainte-Marthe L.,  
 RA Verdier J., Verdier-Discala C., Hiller L.W., Fulton L., McPherson J.,  
 RA Marsuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W.,  
 RA Ouetier F., Waterston R., Hood L., Weissenbach J.,  
 RT "The DNA sequence and analysis of human chromosome 14.";  
 RT Nature 421:601-607(2003).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hesteh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Tshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SIMILARITY: Belongs to the MISS family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AL834285; CAD38959.1; ALT\_INIT.  
 DR EMBL: AL119316; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: BC015621; AAH15621.1; -;  
 DR Genew; HGNC:19840; C14orf32.  
 FT DOMAIN 27 241 PRO-RICH.  
 SQ SEQUENCE 245 AA; 24269 MW; 8B03ADFE581D5C5 CRC64;  
 Query Match 14.6%; Score 198; DB 1; Length 245;  
 Best Local Similarity 39.9%; Pred. No. 5.9e-06;  
 Matches 57; Conservative 5; Mismatches 47; Indels 34; Gaps 9;  
 QY 1 MANDFSIHDAL-----SGSGNPNPQGMPGA--WGNDPAGAGGYPGAP 45  
 DB 1 MSDFSLADALPEHSAPKTSVSNTPKQGP-PQGMPSGNPNWNPDSA-----PSSVPSGLP 54





```

DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD000007; C1g helix; 3.
DR PROSITE; PS01208; VWF_C_1; PARTIAL.
KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KM Glycoprotein; Collagen.
FT DOMAIN 1 14
FT DOMAIN 15 1040
FT DOMAIN 1041 1049
FT MOD_RES 95 95
FT MOD_RES 107 107
FT MOD_RES 119 119
FT MOD_RES 938 938
FT MOD_RES 950 950
FT CARBOHYD 107 107
FT CARBOHYD 950 950
FT DISULFID 1040 1040
FT DISULFID 1041 1041
SQ SEQUENCE 1049 AA; 93651 MW; 8EEC3D1C66BC9A3 CRC64;

Query Match 14.2%; Score 193; DB 1; Length 1049;
Best Local Similarity 33.7%; Pred. No. 4.5e-05;
Matches 64; Conservative 8; Mismatches 58; Indels 60; Gaps 12;

QY 5 FSLHDLG-----SGNNPQGWPGAMGNQAGAGYFGAYPGY--QAPPGAYPGA 57
DB 2 YRAYDVKSSVAGGAGIAGYGPAGPPPG--PGTSGHGA--PGAPGYGPPGE--PGA 56
QY 58 PGAPYHAGAGAY-----PGAPAGYVYGP--SGPGAYP----- 89
DB 57 GPAGPPPGAGCGPSKQDSGSRPGGRGFRPPGCMGPGAMPFPCKMKGHRGDRN 116
QY 90 -SSGQPSAPGAY---ATGPPYAPA--GPLIVPNYLPFGVVPRLITLIGTVKPMNR 142
DB 117 GKGGRGAGLKGNGVPGEDGAPGMRGAPGGRGPRG-----LPGAAGARGN- 166
QY 143 IALDPORGND 152
DB 167 ---DGARGSD 173

RESULT 23
ANX7_MOUSE STANDARD; PRT; 463 AA.
ID ANX7_MOUSE
AC Q07076;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Annexin A7 (Annexin VII) (Synexin).
GN ANXA7 OR ANX7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93168121; PubMed=7916616;
RA Zhang-Xeck Z.Y., Burns A.L., Pollard H.B.;
RT "Mouse synexin (annexin VII) polymorphisms and a phylogenetic
RT comparison with other synexins.";
RL Biochem. J. 289:735-741(1993).
CC -1- FUNCTION: Calcium/phospholipid-binding protein which promotes
CC membrane fusion and is involved in exocytosis.
CC -1- DOMAIN: A pair of annexin repeats may form one binding site for
CC calcium and phospholipid.
CC -1- SIMILARITY: Belongs to the annexin family.
CC -1- SIMILARITY: Contains 4 annexin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; I13129; AAA37238.1; -.
DR HSSP; P26256; 1DM5.
DR MGP; MGI:88031; Anxa7.
DR InterPro; IPR001464; Annexin.
DR Pfam; PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXIN.
DR ProDom; PD000143; Annexin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXIN; 4.
KM Annexin; Calcium/phospholipid-binding; Repeat.
FT DOMAIN 1 143
FT REPEAT 169 229
FT REPEAT 241 301
FT REPEAT 324 384
FT REPEAT 400 460
FT DOMAIN 5 20
FT REPEAT 5 9
FT REPEAT 10 14
FT REPEAT 16 20
SQ SEQUENCE 463 AA; 49939 MW; 50F7B20FD48EB0D5 CRC64;

Query Match 14.2%; Score 192.5; DB 1; Length 463;
Best Local Similarity 33.0%; Pred. No. 2.3e-05;
Matches 74; Conservative 16; Mismatches 77; Indels 57; Gaps 14;

QY 11 ISGSNPNQGWPGAMGNQAG-----AGYFGAYVYGPYGAAP--GAYPGQAAPG 60
DB 1 MSYPPYP-PTGYPPPPGYPAGQSSFTAGY-----YSGPPPMGAGAYP--PAPSG 52
QY 61 AYHAGPAGAY---GAPAGYVYGP--PPSGPAYS-----SGQSSAG-AYA 101
DB 53 GYPGA-GGYPAPGYPAGYVYGPAGALSPGPPAYPGQGGPAGPAGFGYGPQPPAOSYG 111
QY 102 TGPY-----GAPAGPLIVPNYLPFGGVV-----RMLTILGTVKPMNRIMLDFOR 149
DB 112 GGPAGVVPVPGFPFGGQMSQY---PGGAPYPPSQPASMTOGTGTLIPASN---FDAMR 164
QY 150 GNDVAFHPPFENNRRYIVCNTKLDNNMGREROSVFPFESGK 193
DB 165 DAEILRKAMKGRGTDEQALVDVVSNRSDQRIYAAFTWTK 208

RESULT 24
LE32_CAEEL STANDARD; PRT; 279 AA.
ID LE32_CAEEL
AC P36573;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 32 kDa beta-galactoside-binding lectin (32 kDa GBP).
GN W09H1.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92348399; PubMed=1639789;
RA Hirabayashi J., Satoh M., Kasai K.-I.;
RT "Evidence that Caenorhabditis elegans 32-kDa beta-galactoside-binding
RT protein is homologous to vertebrate beta-galactoside-binding lectins.
RT cDNA cloning and deduced amino acid sequence.";
RL J. Biol. Chem. 267:15485-15490(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476274; PubMed=9334250;
RA Arata Y., Hirabayashi J., Kasai K.-I.;
RT "Structure of the 32-kDa galectin gene of the nematode Caenorhabditis

```

RT elegans";  
 RL J. Biol. Chem. 272:26669-26677(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Smye R.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 202-220.  
 RX MEDLINE=92348337; PubMed=1639749;  
 RA Hirabayashi J., Satoh M., Ohnaya Y., Kasai K.-I.;  
 RT "Purification and characterization of beta-galactoside-binding  
 proteins from Caenorhabditis elegans.";  
 RL J. Biochem. 111:553-555(1992).  
 CC -1- FUNCTION: Binds galactose.  
 CC -1- SIMILARITY: Belongs to the galectin (galactin/S-lectin) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M94671; AAB87718.1; -;  
 DR EMBL; AB000802; BAA22942.1; -;  
 DR EMBL; Z82081; CAB04959.1; -;  
 DR PIR; PX0062; PX0062.  
 DR PIR; T37216; T37216.  
 DR HSSP; P17931; 1A3K.  
 DR WormPep; W09H1.6a; CE16576.  
 DR InterPro; IPR008985; Cona\_like\_1ec\_g1.  
 DR InterPro; IPR001079; Galectin.  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.  
 DR SMART; SM00276; GLECT; 2.  
 DR PROSITE; PS00309; GALAPTIN; 2.  
 DR Galectin; Lectin; Repeat; Multigene family.  
 FT DOMAIN 1 146 GALAPTIN 1.  
 FT DOMAIN 147 279 GALAPTIN 2.  
 FT BINDING 213 219 BETA-GALACTOSIDE (BY SIMILARITY).  
 SQ SEQUENCE 279 AA; 31809 MW; ED9AE2A837571DA CRC64;

Query Match 13.9%; Score 188.5; DB 1; Length 279;  
 Best Local Similarity 36.4%; Pred. No. 2.6e-05;  
 Matches 47; Conservative 26; Mismatches 49; Indels 7; Gaps 5;

QY 114 VPYALPLPGVVPRLITITLTKVKNRRIALDFQRGN-DVAHPF-PFENNRRIVVCN 171  
 DB 150 VPYSGLANGLPVGKSLVFGTVEKKARFFHVNILRKNGDISFFHNPFDKX--VLRN 206  
 QY 172 TKLNNWGREEROSVFPESGKPFKIQVLVEPDHPKXAVN-DALHQNHRKXKLNESKL 230  
 DB 207 SLANWENREKGNPFKGVGDLVYQNEEYAVQVFNNGERYISFPHRDP-HDIAGL 265  
 QY 231 GISGDIDLT 239  
 DB 266 QISGDIELS 274

RESULT 25  
 CBPA DICTI STANDARD; PRT; 467 AA.  
 AC P35085;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Calcium-binding protein.  
 GN CBPP OR CBPA.  
 OS Dictyostelium discoideum (Slime mold).  
 CC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.  
 CC NCBL\_TaxID=44689;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RA Wenington R., Greenwood M., Tsang A.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U03413; AAA03471.1; -;  
 DR DictyBase; DDB0002101; cbp.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR006031; XYPPX.  
 DR Pfam; PF00036; efhand; 2.  
 DR Pfam; PF02162; XYPPX; 18.  
 DR SMART; SM00054; EFH; 2.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 DR Calcium-binding.  
 FT CA BIND 412 423 EF\_HAND (POTENTIAL).  
 SQ SEQUENCE 467 AA; 48976 MW; 2D2055D1E344FC18 CRC64;

Query Match 13.9%; Score 188; DB 1; Length 467;  
 Best Local Similarity 37.2%; Pred. No. 4.4e-05;  
 Matches 54; Conservative 0; Mismatches 35; Indels 56; Gaps 8;

QY 12 SGSGNPNGQWPGAMGNOPAGAGGVP-----GASYPGYPGAP-----GAYPGAP 58  
 DB 113 SGQYPPQPGQPGYPPQPGAPGYPQPGQPGQPGQPGQPGQPGQPGQPGQPG 172  
 QY 59 -PGAY-----HGAPAY-----PGAPAP-GVYPPGPPSGP----- 85  
 DB 173 QPGAYPPQPGQPGQPGAYPPQPGQVQNTLAKTGAQPGYPPPGGAYPPGQPGAYPPGQ 232  
 QY 86 ----GAYPSSQPSA-----PGAY 100  
 DB 233 QPPWAGYPPQPGQPGAYPPQPGQPGAY 257

RESULT 26  
 LEG1 HAECO STANDARD; PRT; 283 AA.  
 ID LEG1 HAECO  
 AC 044126;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 32 kDa beta-galactoside-binding lectin (galectin 1).  
 GN GAL-1.  
 OS Haemonchus contortus (Barber pole worm).  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
 CC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
 CC NCBL\_TaxID=6289;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Moredu; TISSUE=Gut;  
 RX MEDLINE=20067137; PubMed=10599080;  
 RA Newlands G.F.J., Skuce P.J., Knox D.P., Smith S.K., Smith W.D.;  
 RT "Cloning and characterization of a beta-galactoside-binding protein  
 RT (galectin) from the gut of the gastrointestinal nematode parasite  
 RT Haemonchus contortus.";  
 RT Parasitology 119:483-490(1999).  
 CC -1- FUNCTION: Binds galactose.  
 CC -1- SIMILARITY: Belongs to the galectin (galaptin/S-lectin) family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way











RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
 RA Hayashizaki Y.;  
 RL "Functional annotation of a full-length mouse cDNA collection.";  
 CC Nature 409:685-690(2001).  
 CC -1- FUNCTION: Major component of the Descemet's membrane (basement  
 CC membrane) of corneal endothelial cells.  
 CC -1- SUBUNIT: May form homotrimers, or heterotrimers in association  
 CC with alpha 2(VIII) type collagens.  
 CC -1- TISSUE SPECIFICITY: High levels in calvarium, eye and skin of  
 CC newborn mice; also in various epithelial, endothelial and  
 CC mesenchymal cells.  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.  
 CC -1- SIMILARITY: Contains 1 C1Q domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X66976; CAA47387.1; -.  
 DR EMBL; X66977; CAA47387.1; JOINED.  
 DR EMBL; AK018742; BAB31383.1; -.  
 DR PIR; S23779; S23779.  
 DR MGI; 88463; Col6a1.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR008983; TNF\_like.  
 DR Pfam; PF00386; C1q; 1.  
 DR DR Pfam; PF01391; Collagen; 7.  
 DR PRINTS; PR00007; COLLENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Cell adhesion; Collagen; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 743 COLLAGEN ALPHA 1(VIII) CHAIN.  
 FT DOMAIN 118 117 NONHELIICAL REGION (NC2).  
 FT DOMAIN 572 743 TRIPLE-HELICAL REGION (COL1).  
 FT DOMAIN 608 743 NONHELIICAL REGION (NC1).  
 FT DOMAIN 5 6 C1Q.  
 FT CONFLICT 85 85 G-> R (IN REF. 2).  
 FT CONFLICT 109 109 H-> Y (IN REF. 1).  
 FT CONFLICT 248 248 K-> KG (IN REF. 2).  
 FT CONFLICT 313 313 P-> L (IN REF. 1).  
 FT CONFLICT 323 324 P-> A (IN REF. 2).  
 FT CONFLICT 361 361 IP-> SR (IN REF. 1).  
 FT CONFLICT 596 596 D-> H (IN REF. 1).  
 FT CONFLICT 717 719 T-> E (IN REF. 2).  
 FT CONFLICT 719 719 MPS-> NP (IN REF. 1).  
 SQ SEQUENCE 743 AA; 73454 MW; F584D85BD53897F4 CRC64;

Query Match 13.5%; Score 183.5; DB 1; Length 743;  
 Best Local Similarity 34.6%; Pred. No. 0.00013;  
 Matches 55; Conservative 8; Mismatches 57; Indels 39; Gaps 8;

QY 15 GNPNPQGMGAMGNQAGAGGYPG-ASYGXYPG-----QAPPGAYPGQA 57  
 DB 447 GPPGWRGLPPIG-PKSGEGHKGJLPGVGLGPKSGPIPGDGLQSPPG-IPGIV 503  
 QY 58 PPGAYHAGP-----AYPGAPGYP-----GPPSGRAYSSQSPAPRAYA 101  
 DB 504 GPSGPIGPGI-GPKSGEPGLGPPGFPPVGRKGVAGLHGPPKPGALGPQGFHFG--P 561  
 QY 102 TGPYGAAPLIVPYNLPLPGGVVPRMLITIGTKPNA 140  
 DB 562 PGPPGPPGPAVMPFPPSP-QGEYLPDMGLGIDGVTPHNA 599

RESULT 31  
 ID LEGC HUMAN STANDARD; PRT; 336 AA.  
 AC Q96DT0; Q96DS9; Q96PR9; Q9H259; Q9H259; Q9NZ02;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Galectin-12 (Galectin-related inhibitor of proliferation).  
 GN LGALS12 OR GRP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP TISSUE FROM N.A. (ISOFORMS A; B; C AND D).  
 RC TISSUE=Adipose tissue;  
 RX MEDLINE=21423969; PubMed=11435439;  
 RA Hotta K., Funahashi T., Matsukawa Y., Takahashi M., Nishizawa H.,  
 RA Kohida K., Matsuda M., Kuriyama H., Kihara S., Nakamura T.,  
 RA Tochino Y., Bodkin N.L., Hansen B.C., Matsuzawa Y.;  
 RT "Galectin-12, an adipose-expressed galectin-like molecule possessing  
 RT apoptosis-inducing activity.";  
 RL J. Biol. Chem. 276:34089-34097(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS B; E AND F).  
 RC TISSUE=Retina;  
 RX MEDLINE=21283005; PubMed=11283015;  
 RA Yang R.-Y., Hsu D.K., Yu L., Ni J., Liu F.-T.;  
 RT "Cell cycle regulation by galectin-12, a new member of the galectin  
 RT superfamily.";  
 RL J. Biol. Chem. 276:20252-20260(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC TISSUE=Blood;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg K.H., Buetow K.H., Schaefer C.F., Blat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueidi T.B., Toshiyuki S., Carrincci P., Prange C.,  
 RA Rana S.S., Loguelli N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W.,  
 RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Matra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Binds lactose.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Name=A;  
 CC IsoId=Q96DT0-1; Sequence=Displayed;  
 CC Name=B; Synonyms=GRP1a;  
 CC IsoId=Q96DT0-2; Sequence=VSP\_003100;  
 CC Name=C;  
 CC IsoId=Q96DT0-3; Sequence=VSP\_003102;  
 CC Name=D;  
 CC IsoId=Q96DT0-4; Sequence=VSP\_003100, VSP\_003102;  
 CC Name=E; Synonyms=1;  
 CC IsoId=Q96DT0-5; Sequence=VSP\_003099;  
 CC Name=F; Synonyms=2;  
 CC IsoId=Q96DT0-6; Sequence=VSP\_003099, VSP\_003101;  
 CC -1- TISSUE SPECIFICITY: Not widely expressed. Predominantly expressed



```
CC12_CAEEL
ID _CC12_CAEEL STANDARD; PRT; 316 AA.
AC P20630;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cuticle collagen 12 precursor.
GN COL-12 OR F15H10.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90172409; PubMed=1689778;
RA Park Y.-S., Kramer J.M.;
RT "Randomly duplicated Caenorhabditis elegans collagen genes differ in
RT their modes of splicing."
RL J. Mol. Biol. 211:395-406(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -1- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -1- SIMILARITY: Belongs to the cuticular collagen family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X51622; CA35954.1; -.
DR EMBL; Z73972; CA98257.1; -.
DR PIR; S08159; S08159.
DR WormPep; F15H10.1; CE05638.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 3.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW Signal.
FT SIGNAL 1 36
FT CHAIN 37 316
FT DOMAIN 128 157 TRIPLE-HELICAL REGION.
FT DOMAIN 176 202 TRIPLE-HELICAL REGION.
FT DOMAIN 206 235 TRIPLE-HELICAL REGION.
FT DOMAIN 240 266 TRIPLE-HELICAL REGION.
FT DOMAIN 269 304 TRIPLE-HELICAL REGION.
SQ SEQUENCE 316 AA; 30098 MW; 6CA81FP94706D42E CRC64;

Query Match 13.2%; Score 179; DB 1; Length 316;
Best Local Similarity 35.5%; Pred. No. 0.00011;
Matches 55; Conservative 4; Mismatches 50; Indels 46; Gaps 8;

QY 13 GSGNPNQGMWPGMGNQPAAGAGYPGASYPGPGQ-----APPG-- 51
Db 126 GSGAGAGAGSPGQDG--APGNDGAPGA--FGNFGQDASEDTAGPDSFCPCDAPGPPGGS 181
QY 52 AAYGQAPPGAYHGAPG-----AYGAPAPSVYPPGPPSGAPYSSGQSPAPGAY-----A 101
Db 182 GAGGQKPPSGAPGAPGAGGAGALFPGPPGAPGPPGAPGAGPAGPAGPAGVVDVPT 241
QY 102 TGPYGAAPGFLIVPYNLP-----LPGGVYPR 127
```

```
Db 242 PGAPGPPGSP--GPAGAPGPGGQAGSSQPGGPPGQ 274
RESULT 34
CC13_CAEEL
ID _CC13_CAEEL STANDARD; PRT; 316 AA.
AC P20631;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cuticle collagen 13 precursor.
GN COL-13 OR F15H10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90172409; PubMed=1689778;
RA Park Y.-S., Kramer J.M.;
RT "Randomly duplicated Caenorhabditis elegans collagen genes differ in
RT their modes of splicing."
RL J. Mol. Biol. 211:395-406(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -1- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -1- SIMILARITY: Belongs to the cuticular collagen family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X51623; CA35955.1; -.
DR EMBL; Z73972; CA98258.1; -.
DR PIR; S08170; S08170.
DR WormPep; F15H10.2; CE05639.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR Pfam; PF01391; Collagen; 3.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW Signal.
FT SIGNAL 1 36
FT CHAIN 37 316
FT DOMAIN 128 157 TRIPLE-HELICAL REGION.
FT DOMAIN 176 202 TRIPLE-HELICAL REGION.
FT DOMAIN 206 235 TRIPLE-HELICAL REGION.
FT DOMAIN 240 266 TRIPLE-HELICAL REGION.
FT DOMAIN 269 304 TRIPLE-HELICAL REGION.
SQ SEQUENCE 316 AA; 30100 MW; 00C6D08BEC4701AF CRC64;

Query Match 13.2%; Score 179; DB 1; Length 316;
Best Local Similarity 35.5%; Pred. No. 0.00011;
Matches 55; Conservative 4; Mismatches 50; Indels 46; Gaps 8;

QY 13 GSGNPNQGMWPGMGNQPAAGAGYPGASYPGPGQ-----APPG-- 51
Db 126 GSGAGAGAGSPGQDG--APGNDGAPGA--FGNFGQDASEDTAGPDSFCPCDAPGPPGGS 181
```











```

RX MEDLINE=98362649; PubMed=9697310;
RA Magnaldo T., Fowles D., Darmon M.;
RT "Galectin-7, a marker of all types of stratified epithelia.";
RL Differentiation 63:159-168(1998).
CC
CC -1- FUNCTION: Could be involved in cell-cell and/or cell-matrix
CC interactions necessary for normal growth control. Pro-apoptotic
CC protein that functions intracellularly upstream of JNK activation
CC and cytochrome c release (By similarity).
CC
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear. May be secreted by
CC a non-classical secretory pathway (By similarity).
CC
CC -1- SIMILARITY: Belongs to the galectin (galactin/S-lectin) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: AF038562; AAB92566.1; -.
CC HSBP: P47929; 2GAL.
CC MGD: MGI:1316742; Lgal57.
CC
CC InterPro: IPR008985; CONA_like_lect_gf.
CC
CC InterPro: IPR001079; Galectin.
CC Pfam: PF00337; Gal-bind_lectin; 1.
CC SMART: SM00276; GLECT; 1.
CC PROSITE: PS00309; GALACTIN; 1.
CC
CC Galectin: Lectin; Apoptosis; Nuclear protein.
CC
CC INT MET 0 BY SIMILARITY.
CC
CC BINDING 69 75 BETA-GALACTOSIDE (POTENTIAL).
CC
CC SEQUENCE 135 AA; 15042 MW; 9C8E3CBF36C5013 CRC64;
CC
CC
CC Query Match 12.7%; Score 173; DB 1; Length 135;
CC Best Local Similarity 33.1%; Pred. No. 0.00012;
CC Matches 42; Conservative 26; Mismatches 49; Indels 10; Gaps 5;
CC
CC QY 120 LPQGVVRMLITLIGTVKPNANRIALDF---QRGNDAVAFH-PRENNRRVIVCTKTL 174
CC ID CA39 CHICK STANDARD; PRT; 675 AA.
CC AC P32017;
CC
CC DT 01-JUL-1993 (Rel. 26, Created)
CC DT 01-JUL-1993 (Rel. 26, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Collagen alpha 3 (IX) chain precursor.
CC GN COL9A3.
CC OS Gallus gallus (Chicken).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC OC Gallus.
CC NCBI_TaxID=9031;
CC
CC RN [1]
CC
CC RP SEQUENCE FROM N.A.
CC
CC RX MEDLINE=92250566; PubMed=1577778;
CC RA Har-El R., Shatma Y.D., Aguilera A., Ueyama N., Wu J.-J.,
CC RA Eyre D.R., Juricic L., Chandrasekaran S., Li M., Nah H.D.,
CC RA Upholt W.B., Tanzer M.L.;
CC RT "Cloning and developmental expression of the alpha 3 chain of chicken
CC type IX collagen.";

```

```

RL J. Biol. Chem. 267:10070-10076(1992).
CC
CC -1- FUNCTION: COLLAGEN TYPE IX IS A MINOR CARTILAGE NONFIBRILLAR
CC COLLAGEN. IT IS ASSOCIATED WITH TYPE II COLLAGEN FIBRILS.
CC
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(IX),
CC ALPHA 2(IX), AND ALPHA 3(IX).
CC
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: M83179; AAB59960.1; -.
CC
CC InterPro: IPR008161; Clg_helix.
CC
CC InterPro: IPR008160; Collagen.
CC Pfam: PF01391; Collagen; 11.
CC
CC ProDom: PD000007; Clg_helix; 3.
CC
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC
CC KW Glycoprotein; Collagen; Signal.
CC
CC FT SIGNAL 1 21
CC FT CHAIN 22 675 COLLAGEN ALPHA 3(IX) CHAIN.
CC FT DOMAIN 25 515 TRIPLE-HELICAL REGION 3 (COL3).
CC FT DOMAIN 516 546 NONHELICAL REGION 3 (NC3).
CC FT DOMAIN 547 626 TRIPLE-HELICAL REGION 2 (COL2).
CC FT DOMAIN 627 631 NONHELICAL REGION 2 (NC2).
CC FT DOMAIN 632 658 TRIPLE-HELICAL REGION 1 (COL1).
CC FT DOMAIN 659 675 NONHELICAL REGION 1 (NC1).
CC FT SITE 242 244 CELL ATTACHMENT SITE (POTENTIAL).
CC FT SITE 591 593 CELL ATTACHMENT SITE (POTENTIAL).
CC FT CARBOHYD 479 479 N-LINKED (GLCNAC...) (POTENTIAL).
CC
CC SEQUENCE 675 AA; 63013 MW; C983FBC924A10098 CRC64;
CC
CC
CC Query Match 12.7%; Score 173; DB 1; Length 675;
CC Best Local Similarity 32.2%; Pred. No. 0.00052;
CC Matches 59; Conservative 14; Mismatches 64; Indels 46; Gaps 11;
CC
CC QY 9 DALSG-GNPNFGMGWPGMGN-CPAGAGYPGASYPGQAPGAYPG---CAPPGAY 62
CC ID CA13 MOUSE STANDARD; PRT; 1464 AA.
CC AC P08121; Q61429; Q9CNR7;
CC
CC DT 01-AUG-1988 (Rel. 08, Created)
CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Collagen alpha 1 (III) chain precursor.
CC GN COL3A1.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
CC
CC RN [1]
CC
CC RP SEQUENCE FROM N.A.
CC
CC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;

```

FX MEDLINE=95011609; PubMed=7926795;  
 RA Toman D., de Crombrughe B.;  
 RT "The mouse type-III procollagen-encoding gene: genomic cloning and  
 RT complete DNA sequence.";  
 RL Gene 147:161-168(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang Y., Hsien P.,  
 RA Datchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stempleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rask S.S., Loguélus N.A., Peters G.J., Abramson R.D., Mollath S.J.,  
 RA Roark S.A., McKernan P.J., McKernan K.J., Malek J.A., Gnartnre P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywninski M.I., Skalka U., Smalls D.E.,  
 RA Schermer A., Schein U.E., Jones S.U.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 1-488 FROM N.A.  
 RC MEDLINE=88167859; PubMed=2443309;  
 RX Wood U., Theriault N., Vogel G.;  
 RA "Complete nucleotide sequence of the N-terminal domains of the murine  
 RT alpha-1 type-III collagen chain.";  
 RL Gene 61:225-230(1987).  
 RN [4]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RC MEDLINE=85131189; PubMed=3972847;  
 RX Lian G., Mudryj M., de Crombrughe B.;  
 RA "Identification of the promoter and first exon of the mouse alpha 1  
 RT (III) collagen gene.";  
 RL J. Biol. Chem. 260:3773-3777(1985).  
 RN [5]  
 RP SEQUENCE OF 810-1464 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai U., Shingawa A.A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Akaiwa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aikawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Pletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Knebel P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Maehio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bartz G.,  
 RA Blake U., Boffelli D., Bojuna R., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guelinich S., Hill D., Hofmann M., Hune D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli C., Monodarts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakomoto N.,  
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Yashnawa-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP SEQUENCE OF 1442-1464 FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=91274355; PubMed=2054384;  
 RA Metcaterana M., Roman D., de Crombrughe B., Vucorio E.;  
 RT "Specific hybridization probes for mouse type I, II, III and IX

```

RT collagen mRNAs." ;
RL Biochim. Biophys. Acta 1089:241-243(1991) .
CC -I- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -I- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxyllysines.
CC -I- PM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains.
CC -I- PM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC the oxygen atom of a post-translationally added hydroxy[ group (by
CC similarity). Contains 1 WFEC domain.
CC -I- SIMILARITY: Contains 1 WFEC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage fee and for commercial/
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52046; CAA36279.1; -.
CC DR EMBL; BC03089; AAH3089.1; -.
CC DR EMBL; M858724; AHS58724.1; -.
CC DR EMBL; M18933; AAA37338.1; -.
CC DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
CC DR EMBL; AK019448; BAB31724.1; -.
CC DR EMBL; X57983; CAA41048.1; -.
CC DR PIR; A27353; A27353.
CC DR PIR; S59856; S59856.
CC DR MED; MG1:88453; Col3a1.
CC DR InterPro; IPR008161; C1g_helix.
CC DR InterPro; IPR008160; Collagen.
CC DR InterPro; IPR000885; F1b_collagen_C.
CC DR InterPro; IPR002181; Fibrinogen_C.
CC DR InterPro; IPR01007; VWF_C.
CC DR Pfam; PF01410; COLF1; 1.
CC DR Pfam; PF01391; Collagen; 18.
CC DR ProDom; PD000007; C1g_helix; 1.
CC DR ProDom; PD002078; F1b_collagen_C; 1.
CC DR SMART; SMO0038; COLF1; 1.
CC DR SMART; SMO0214; VMC; 1.
CC DR PROSITE; PS01208; WFEC_1; 1.
CC DR PROSITE; PS50184; VWFEC_2; 1.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Signal.
CC MW SIGNAL
CC FT SIGNAL 1 23
CC FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
CC FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
CC FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
CC FT DOMAIN 31 90 WFEC.
CC FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).
CC FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
CC FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).
CC FT CARBOHD 262 262 O-LINKED (GAL. ...) (BY SIMILARITY).
CC FT MOD RES 262 262 HYDROXYLATION (BY SIMILARITY).
CC FT MOD RES 283 283 HYDROXYLATION (BY SIMILARITY).
CC FT MOD RES 859 859 HYDROXYLATION (BY SIMILARITY).
CC FT MOD RES 976 976 HYDROXYLATION (BY SIMILARITY).
CC FT MOD RES 1093 1093 HYDROXYLATION (BY SIMILARITY).
CC FT MOD RES 1105 1105 HYDROXYLATION (BY SIMILARITY).
CC FT DISULFD 1195 1195 INTERCHAIN (BY SIMILARITY).
CC FT DISULFD 1196 1196 INTERCHAIN (BY SIMILARITY).
CC SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

Query Match 12.7%; Score 172.5; DB 1; Length 1464;
Best Local Similarity 32.5%; Pred. 0.0011;
Matches 62; Conservativity 7; Mismatches 61; Indels 61; Gaps 12;

```

```

Db 156 FDSYVKGSGVGMGTPGPBAGPPPG--PGSSGHPGS--PGSPGYGPPGE-PGQAP 210
QY 60 GAYHAGPAGAY-----PGAPAPGVYPPPP--SGPGAYP----- 89
Db 211 AGPPPPALGSPAGAGKDGSGRGRORRORRCLPPPPGKGPAGPFGPKMGHGFDR 270
QY 90 --SSGQPSAPGAYA---TGPYGA PA--GPLLVPYNLP.PGSGVPRMTITLGVKPNAN 141
Db 271 NGEKGTGAPGLKGENGLPGDNGALGPMQPRGAPGERGRPG-----LFGAGARGN 321
QY 142 RIALDFGRND 152
Db 322 ---DGARGSD 328

RESULT 43
CC02 CAEEL
ID _CC02 CAEEL STANDARD; PRT; 301 AA.
AC PI7656;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cuticle collagen 2 precursor.
GN COL-2 OR M01B6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=83050944; PubMed=7139711;
RA Kramer J.M., Cox G.N., Hirsch D.;
RT "Comparisons of the complete sequences of two collagen genes from
RT Caenorhabditis elegans.";
RL Cell 30:599-606 (1982).
EN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=85105075; PubMed=2578467;
RA Kramer J.M., Cox G.N., Hirsch D.;
RT "Expression of the Caenorhabditis elegans collagen genes col-1 and
RT col-2 is developmentally regulated.";
RL J. Biol. Chem. 260:1945-1951 (1985).
RN [3]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC STRAIN=Bristol N2;
RA Kohara Y., Shin-I T., Suzuki Y., Sugano S., Thierry-Mieg D.,
RT "The Caenorhabditis elegans transcriptome project, a complementary
RT view of the genome.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Coles L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -1- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -1- TISSUE SPECIFICITY: Syncytial dorsal and ventral epidermis.
CC -1- DEVELOPMENTAL STAGE: Low level expression from embryos at comma
CC stage to adult; maximal level in dater larvae.
CC -1- SIMILARITY: Belongs to the cuticular collagen family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01048; AAA27990.1; -.
DR EMBL; V00148; CAA23464.1; -.
DR EMBL; AY347575; AAP97869.1; -.
DR EMBL; Z68301; CAA92620.1; -.
DR PIR; B31219; B31219.
DR WormPep; M01B6.7; CE03759.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR Pfam; PF01391; Collagen; 2.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW Signal.
FT SIGNAL. 1 37
FT CHAIN 38 301 POTENTIAL.
FT DOMAIN 85 97 CUTICLE COLLAGEN 2.
FT DOMAIN 105 134 GLY-RICH.
FT DOMAIN 153 176 TRIPLE-HELICAL REGION.
FT DOMAIN 183 212 TRIPLE-HELICAL REGION.
FT DOMAIN 215 282 TRIPLE-HELICAL REGION.
SQ SEQUENCE 301 AA; 28025 MW; 33317E3BDAC302F9 CRC64;

Query Match 12.7%; Score 172; DB 1; Length 301;
Best Local Similarity 39.3%; Pred. No. 0.00029;
Matches 46; Conservative 3; Mismatches 60; Indels 8; Gaps 3;

QY 12 SGGNPNDQGWPGAWGNPA---GAGYPGASYPGYGPAGPAGYAGQAPPGAYHAGPG 67
Db 180 SGPPAGPPGPGAGAGNDGAPGAPGPGEPGASEGSGGEPGPPGPPGAGNDGAPG 239
QY 68 AYPGAPAGVYPPGPPSGGSPGAPGAYATGPGAPAGPLIVPNLPLPGCV 124
Db 240 T--GGPDPAGPKGPPGAAGAPADONPGPG--TAGKPGGEGKCTCKYCAIDGV 292

RESULT 44
SSA2 HUMAN
ID _SSA2 HUMAN STANDARD; PRT; 464 AA.
AC Q15428; 075245;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Splicing factor 3A subunit 2 (Spliceosome associated protein 62) (SAP
DE 62) (SF3A66).
GN SF3A2 OR SAP62.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94023929; PubMed=8211113;
RA Bennett M., Reed R.;
RT "Correspondence between a mammalian spliceosome component and an
RT essential yeast splicing factor.";
RL Science 262:105-108 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RC Burkhardt-Schultz K.J., Gordon L., Kyle A., Ramirez M., Seilwagen S.,
RC Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RC Dangaran L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RC Liu S., Attix C., Andreise T., Frankheim M., Amico-Keller G.,
RC Cofield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RC Kronmiller B., Arellano A.S., Montgomery M., Ow D., Nolan M., Trong S.,
RC Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.5 kb contig in human 19p13.3 containing a
RT serine protease gene cluster";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```

RC TISSUE=lung, and lymph;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Strausberg R.D., Collins F.S., Wagner I., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.B., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.O., Ustin J.B., Toshlyuk S., Carninci P., Prange C.,  
 RA Bask S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Rodrikuez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP CHARACTERIZATION OF THE SPLICEOSOME.  
 RX MEDLINE=20337962; PubMed=10882114;  
 RA Das R., Zhou Z., Reed R.;  
 RT "Functional association of U2 snRNP with the ATP-independent  
 RT spliceosomal complex E.";  
 RL Mol. Cell 5:779-787(2000).  
 CC -1- FUNCTION: Subunit of the splicing factor SF3A required for 'A'  
 CC complex assembly formed by the stable binding of U2 snRNP to the  
 CC branchpoint sequence (BPS) in pre-mRNA. Sequence independent  
 CC binding of SF3A/SF3B complex upstream of the branch site is  
 CC essential, it may anchor U2 snRNP to the pre-mRNA. May also be  
 CC involved in the assembly of the 'E' complex.  
 CC -1- SUBUNIT: COMPONENT OF SPLICING FACTOR SF3A WHICH IS COMPOSED OF  
 CC THREE SUBUNITS: SF3A3/SAP61, SF3A2/SAP62, SF3A1/SAP14. SF3A  
 CC ASSOCIATES WITH THE SPLICING FACTOR SF3B AND A 12S RNA UNIT TO  
 CC FORM THE U2 SMALL NUCLEAR RIBONUCLEOPROTEIN COMPLEX (U2 snRNP).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SF3A2 FAMILY.  
 CC -1- SIMILARITY: Contains 1 matrix-type zinc finger.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: L21990; AAA60301.1; -;  
 DR EMBL: AC005263; AAC25613.1; -;  
 DR EMBL: BC004434; AAH04434.1; -;  
 DR EMBL: BC009903; AAH09903.1; -;  
 DR PIR: A47655; A47655.  
 DR Genew: HGNC:10766; SP3A2.  
 DR CK: Q15428; -;  
 DR MIM: 600796; -;  
 DR GO: GO:0030532; C:small nuclear ribonucleoprotein complex; TAS.  
 DR GO: GO:0005681; C:spliceosome complex; TAS.  
 DR GO: GO:0006371; P:mRNA splicing; TAS.  
 DR InterPro: IPR007087; Znf\_C2H2.  
 DR InterPro: IPR000690; Znf\_matriin.  
 DR InterPro: IPR003604; Znf\_U1.  
 DR SMART: SM00355; Znf\_C2H2; 1.  
 DR SMART: SM00451; Znf\_U1; 1.  
 DR PROSITE: PS00171; ZF\_MATRIN; 1.  
 KW Spliceosome; mRNA processing; mRNA splicing; Nuclear protein;  
 KW Zinc-finger; Repeat.  
 FT ZN\_FING 54 84 MATRIN-TYPE.  
 FT DOMAIN 233 236 POLY-PRO.  
 FT DOMAIN 252 256 POLY-PRO.

FT DOMAIN 458 462 POLY-PRO.  
 FT CONFLICT 29 29 R -> P (IN REF. 1).  
 SQ SEQUENCE 464 AA; 49255 MW; FA46F064A55EA2CB CXC64;  
 Query Match 12.6%; Score 171.5; DB 1; Length 464;  
 Best Local Similarity 37.4%; Pred. No. 0.00046;  
 Matches 49; Conservative 11; Mismatches 52; Indels 19; Gaps 7;  
 QY 7 LHDALSSGNNPQGMNQPAGAGYRCAS--YGYGQAPPPAGAP--PGAY 62  
 DB 224 VHPASGVHPAPGVHPAPGVHPAPGVHPPTSGVHPAPGVHPA--PGVHPAPGVH 351  
 QY 63 HCAPAYGAPAPGVYPCP----PSCGAYPSSG--QPSAPGAYATGPGYCAPLIVP 115  
 DB 352 PPAPGVHP--PAPGVHPSPAGVHPQAPGVHPAALVHPAPGVHPAPGVHPAPGVHP 409  
 QY 116 YNLPLPGGVVP 126  
 DB 410 Q----PPGVHP 416  
 RESULT 45  
 CA24\_HUMAN  
 ID CA24\_HUMAN STANDARD; PRT; 1712 AA.  
 AC P08572;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 GN COL4A2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89066769; PubMed=3198637;  
 RA Hostikka S.L., Tryggvason K.;  
 RT "The complete primary structure of the alpha 2 chain of human type IV  
 RT collagen and comparison with the alpha 1(IV) chain.";  
 RL J. Biol. Chem. 263:19488-19493(1988).  
 RN [2]  
 RP SEQUENCE OF 1-1042 FROM N.A.  
 RC TISSUE=placenta;  
 RX MEDLINE=88151998; PubMed=3345760;  
 RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;  
 RT "Human basement membrane collagen (type IV). The amino acid sequence  
 RT of the alpha 2(IV) chain and its comparison with the alpha 1(IV)  
 RT chain reveals deletions in the alpha 1(IV) chain.";  
 RL Eur. J. Biochem. 172:35-42(1988).  
 RN [3]  
 RP SEQUENCE OF 1254-1712 FROM N.A.  
 RX MEDLINE=87219158; PubMed=3582677;  
 RA Hostikka S.L., Kurkinen M., Tryggvason K.;  
 RT "Nucleotide sequence coding for the human type IV collagen alpha 2  
 RT chain cDNA reveals extensive homology with the NC-1 domain of alpha 1  
 RT (IV) but not with the collagenous domain or 3'-untranslated region.";  
 RL FEBS Lett. 216:281-286(1987).  
 RN [4]  
 RP SEQUENCE OF 1451-1485 FROM N.A.  
 RX MEDLINE=87092438; PubMed=3025878;  
 RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;  
 RT "Human collagen genes encoding basement membrane alpha 1 (IV) and  
 RT alpha 2 (IV) chains map to the distal long arm of chromosome 13.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).  
 RN [5]  
 RP SEQUENCE OF 1486-1712 FROM N.A.  
 RX MEDLINE=87250571; PubMed=2449508;  
 RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;  
 RT "Duplication of type IV collagen COOH-terminal repeats and species-  
 RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";  
 RL J. Biol. Chem. 262:9231-9238(1987).  
 RN [6]

RP SEQUENCE OF 1-33 FROM N.A.  
 RA MEDLINE=89034231; PubMed=3182844;  
 RX Solimena R., Huotari M., Hostilka S.L., Prockop D.J., Trygvaason K.;  
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
 RT collagen are divergently encoded on opposite DNA strands and have an  
 RT overlapping promoter region."  
 RL J. Biol. Chem. 263:1721-17220(1988).  
 RN [17]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=89030632; PubMed=2846280;  
 RA Poeschl E., Pollner R., Kuehn K.;  
 RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human  
 RT basement membrane collagen type IV are arranged head-to-head and  
 RT separated by a bidirectional promoter of unique structure."  
 RL EMBO J. 7:2687-2695(1988).  
 RN [8]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=93305049; PubMed=8317999;  
 RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;  
 RT "Identification of a novel sequence element in the common promoter  
 RT region of human collagen type IV genes, involved in the regulation of  
 RT divergent transcription."  
 RL Biochem. J. 292:687-695(1993).  
 RN [9]  
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.  
 RC TISSUE=Placenta;  
 RX MEDLINE=89005112; PubMed=2844531;  
 RA Siebold B., Deutmann R., Kuehn K.;  
 RT "The arrangement of intra- and intermolecular disulfide bonds in the  
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
 RT of basement-membrane type IV collagen."  
 RL Eur. J. Biochem. 176:617-624(1988).  
 CC -!- FUNCTION: Type IV collagen is the major structural component of  
 CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
 CC network together with laminins, proteoglycans and entactin/  
 CC nidogen.  
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -  
 CC alpha 6(IV), each of which can form a triple helix structure  
 CC with 2 other chains to generate type IV collagen network.  
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
 CC domain (NC1) at their C-terminus, frequent interruptions of the  
 CC G-X-Y repeats in the long central triple-helical domain (which may  
 CC cause flexibility in the triple helix), and a short N-terminal  
 CC triple-helical 7S domain.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which  
 CC are involved in inter- and intramolecular disulfide bonding. 12 of  
 CC these, located in the NC1 domain, are conserved in all known type  
 CC IV collagens.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X05562; CAA29076.1; -;  
 DR EMBL; X05610; CAA29098.1; -;  
 DR EMBL; J02760; AAS58422.1; -;  
 DR EMBL; M36963; AAS3099.1; -;  
 DR EMBL; X12784; CAA31275.1; -;  
 DR EMBL; J04217; AAS3097.1; -;  
 DR PIR; A32024; CGH02B.  
 DR Genew; HGNC:2203; COL4A2.  
 DR MIM; 120090; -;  
 DR GO; GO:0005587; C:collagen type IV; TAS.  
 DR GO; GO:0005201; P:extracellular matrix structural constituent; TAS.  
 DR GO; GO:0030198; F:extracellular matrix organization and bioge. . .; NAS.  
 DR InterPro; IPR008161; C1g\_helix.

DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 24.  
 DR ProDom; PD000007; C1g\_helix; 7.  
 DR ProDom; PD003923; Procollagen4; 1.  
 DR SMART; SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Signal.  
 FT SIGNAL 1 25  
 FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).  
 FT DISULFD 1504 1593 OR 1590 (BY SIMILARITY).  
 FT DISULFD 1537 1590 OR 1593 (BY SIMILARITY).  
 FT DISULFD 1549 1555 BY SIMILARITY.  
 FT DISULFD 1612 1708 OR 1705 (BY SIMILARITY).  
 FT DISULFD 1646 1705 OR 1708 (BY SIMILARITY).  
 FT DISULFD 1658 1665 BY SIMILARITY.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).  
 FT CONFLICT 471 471 R -> P (IN REF. 2).  
 FT CONFLICT 683 683 A -> G (IN REF. 2).  
 FT CONFLICT 1575 1575 M -> I (IN REF. 5).  
 FT CONFLICT 1663 1663 G -> H (IN REF. 9).  
 FT CONFLICT 1701 1701 H -> G (IN REF. 9).  
 SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;  
 Query Match 12.6%; Score 171.5; DB 1; Length 1712;  
 Best Local Similarity 31.1%; Pred. No. 0.0015;  
 Matches 51; Conservative 15; Mismatches 55; Indels 43; Gaps 7;  
 QY 15 GNPNPQGMPPGAWGNQAPGAGGYRGA-----GYP---GYPGQ-----APP 50  
 DB 1200 GLPPTGKPPSGSDIHEDPEPFPBGRPDGANTLPFGVGPQKGDGAPGERGPP 1259  
 QY 51 GA-----YPGAPPPGAYHGAAPGAPAPGVY-----PGPPSG--PGAYPSSGQP 94  
 DB 1260 GSPGLQSPPTGTPPSSNLSGAG--DKGAPDIFPLKGYRGPFGSALPLGSKDITNP 1316  
 QY 95 SAPGAYVATGPGYGAAPGLVYPYNLPDPGVVPRMLITLLGTVPK 138  
 DB 1317 GAPPTGPKWAGDSGPGRGVPLGPEKGPBGEGMGNTGP 1360  
 RESULT 46  
 ID C413\_HUMAN STANDARD; PRT; 1466 AA.  
 AC P02461; Q15112;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen alpha 1(III) chain precursor.  
 GN COL3A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBT\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin fibroblast;  
 RX MEDLINE=89350838; PubMed=2764886;  
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kulvaneniemi H.,  
 RA Prockop D.J.;  
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
 RT chain of human type III procollagen. Differences in protein structure  
 RT from type I procollagen and conservation of codon preferences."  
 RL Biochem. J. 260:509-516 (1989).  
 RN [2]  
 RP SEQUENCE OF 149-1225 FROM N.A.  
 RX MEDLINE=89386015; PubMed=2780304;  
 RA Janeczko R.A., Ramirez F.;  
 RT "Nucleotide and amino acid sequences of the entire human alpha 1

RT (III) collagen.";  
 RL Nucleic Acids Res. 17:6742-6742(1989).  
 RN [3]  
 RP SEQUENCE OF 168-398.  
 RX MEDLINE=77134724; PubMed=557335;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of cyanogen  
 bromide peptides from the amino-terminal segment of type III collagen  
 of human liver.";  
 RL Biochemistry 16:1158-1164(1977).  
 RN [4]  
 RP REVISIONS.  
 RA Seyer J.M.;  
 RL Submitted (DEC-1977) to the PIR data bank.  
 RN [5]  
 RP SEQUENCE OF 399-727.  
 RX MEDLINE=79000343; PubMed=687591;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of five  
 consecutive CNBr peptides from type III collagen of human liver.";  
 RL Biochemistry 17:3404-3411(1978).  
 RN [6]  
 RP SEQUENCE OF 728-964.  
 RX MEDLINE=80198282; PubMed=6246925;  
 RA Seyer J.M., Mainardi C., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha 1  
 (III)-CB5 from type III collagen of human liver.";  
 RL Biochemistry 19:1583-1589(1980).  
 RN [7]  
 RP SEQUENCE OF 950-1466 FROM N.A.  
 RX MEDLINE=88189827; PubMed=3357782;  
 RA Mankoo B.S., Dalgleish R.;  
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
 RL Nucleic Acids Res. 16:2337-2337(1988).  
 RN [8]  
 RP REVISION TO 1184.  
 RX MEDLINE=89098346; PubMed=3211760;  
 RA Molynieux K., Dalgleish R.;  
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
 RL Nucleic Acids Res. 16:11833-11833(1988).  
 RN [9]  
 RP SEQUENCE OF 1065-1466 FROM N.A.  
 RX MEDLINE=85087944; PubMed=609827;  
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,  
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III  
 procollagen.";  
 RL Nucleic Acids Res. 12:9383-9394(1984).  
 RN [10]  
 RP SEQUENCE OF 965-1200.  
 RX MEDLINE=81208139; PubMed=7016180;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha  
 1(III)-CB9 from type III collagen of human liver.";  
 RL Biochemistry 20:2621-2627(1981).  
 RN [11]  
 RP SEQUENCE OF 1176-1466 FROM N.A.  
 RX MEDLINE=85157600; PubMed=2579949;  
 RA Chu M.-L., Weil D., de Wet M.J., Bernard M.P., Sippola M., Ramirez F.,  
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
 (III) collagen. Partial characterization of the 3' end region of the  
 gene.";  
 RL J. Biol. Chem. 260:4357-4363(1985).  
 RN [12]  
 RP SEQUENCE OF 1161-1200 FROM N.A.  
 RX MEDLINE=86187804; PubMed=3754462;  
 RA Miskulin M., Dalgleish R., Kluever-Beckerman B., Remard S.I.,  
 RT Tolstouhev P., Brantly M., Crystal R.G.;  
 RT "Human type III collagen gene expression is coordinately modulated  
 with the type I collagen genes during fibroblast growth.";  
 RL Biochemistry 25:1408-1413(1986).  
 RN [13]  
 RP SEQUENCE OF 1-170 FROM N.A.

RC TISSUE=Placenta;  
 RX MEDLINE=88303360; PubMed=3405773;  
 RA Toman D., Ricca G., de Crombrughe B.;  
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
 of human prepro alpha 1(III) collagen.";  
 RL Nucleic Acids Res. 16:7201-7201(1988).  
 RN [14]  
 RP SEQUENCE OF 1-176 FROM N.A.  
 RX MEDLINE=89378752; PubMed=2777083;  
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
 RT "Cloning and analysis of the 5' portion of the human type-III  
 procollagen gene (COL3A1).";  
 RL Gene 78:255-265(1989).  
 RN [15]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97259599; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 associated collagen (type IX), and network-forming collagen (type X)  
 cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [16]  
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
 RX MEDLINE=93293988; PubMed=8514866;  
 RA Tromp G., Wu Y., Prockop D.J., Madhatter S.L., Kleinert C.,  
 RA Farley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,  
 RA Cole C.W., Jaakkola P., Rymaszewski M., Pearce W.H., Yao J.S.T.,  
 RA Majumdar K., Smolens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;  
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
 in the triple-helical domain of type III procollagen are an  
 infrequent cause of aortic aneurysms.";  
 RL J. Clin. Invest. 91:2539-2545(1993).  
 RN [17]  
 RP VARIANT THR-698.  
 RX MEDLINE=91045136; PubMed=2235526;  
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusari S.,  
 RA Wu Y., Ganguly A., Prockop D.J.;  
 RT "Wu to A polymorphism in exon 31 of the COL3A1 gene.";  
 RL Nucleic Acids Res. 18:6180-6180(1990).  
 RN [18]  
 RP VARIANT AORTIC ANEURYSM ARG-786.  
 RX MEDLINE=91056145; PubMed=2243125;  
 RA Kontusari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
 with aortic aneurysms.";  
 RL J. Clin. Invest. 86:1465-1473(1990).  
 RN [19]  
 RP VARIANT EDS-IV ARG-828.  
 RX MEDLINE=94016385; PubMed=8411057;  
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
 RT "The substitution of glycine 661 by arginine in type III collagen  
 produces mutant molecules with different thermal stabilities and  
 causes Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 30:690-693(1993).  
 RN [20]  
 RP VARIANT EDS-IV SER-957.  
 RX MEDLINE=89109135; PubMed=2492273;  
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
 RT "A single base mutation that substitutes serine for glycine 790 of  
 the alpha 1 (III) chain of type III procollagen exposes an arginine  
 and causes Ehlers-Danlos syndrome type IV.";  
 RL J. Biol. Chem. 264:1349-1352(1989).  
 RN [21]  
 RP VARIANT EDS-IV VAL-960.  
 RX MEDLINE=95268429; PubMed=7749417;  
 RA Tromp G., de Paeppe A., Nuytink L., Madhatter S.L., Kuivaniemi H.,  
 RT "Substitution of valine for glycine 793 in type III procollagen in  
 Ehlers-Danlos syndrome type IV.";  
 RL Hum. Mutat. 5:179-181(1995).  
 RN [22]  
 RP VARIANT EDS-IV GLU-1014.  
 RX MEDLINE=92316511; PubMed=1352273;

Query Match	12.5%	Score 170;	DB 1;	Length 1466;
Best Local Similarity	33.6%	Pred. No. 0.0016;		
Matches	49;	Conservative	8;	Mismatches 47;
			Indels	42;
			Gaps	7;
QY	15	GNPNPQ-----GWPAGWGNQ-----FAGAGGPGAS-----YRGPYGAAPGAYP	54	
Db	441	GEPPRBERGRGAGTIPVPGAKGDEGDKGSPSEPPGANGIPGAGGERGADGFGPPGPNCTIP	500	
QY	55	GQAPPGAYHGAPG-----AYPGAPAPGVYPPGPPSGPGAYPSSGGQPSAGAY	100	
Db	501	GEKPPAGERGAPGAPGPRGAGGEPGRDGVPPGAPMRGNPSPGSGP---SNGKGPRESQ	557	
QY	101	A-----TGPYGAAGPLVPPNPLPG	122	
Db	558	GESGRPGPPG-PSGPRGQPGVMGPPG	582	

RA Bird D.M.: "Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34  
RT genes; and their deduced collagen products.",  
RL Gene 120(261-266(1992)).

CC -I- FUNCTION: Nematode cuticles are composed largely of collagen-like  
CC proteins. The cuticle functions both as an exoskeleton and as a  
CC barrier to protect the worm from its environment. Mutations in  
CC dpy-13 affects the body shape.  
CC -I- SUBUNIT: Collagen polypeptide chains are complexed within the  
CC unit by disulfide bonds and other types of covalent cross-  
CC links.

CC -I- SIMILARITY: Belongs to the cuticular collagen family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M25559; AAA27994.1; -;  
DR EMBL; U42437; AAA83499.1; -;  
DR PIR; A31921; A31921.  
DR Wormpep; F30B5.1; CE04462.  
DR InterPro; IPR002486; Col cuticle\_N.  
DR InterPro; IPR008160; Collagen.  
DR Pfam; PF01484; Col cuticle\_N.1.  
DR Pfam; PF01391; Collagen; 2.  
DM Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
FT DOMAIN 106 135 TRIPLE-HELICAL REGION.  
FT DOMAIN 154 210 TRIPLE-HELICAL REGION.  
FT DOMAIN 219 278 TRIPLE-HELICAL REGION.  
FT VARIANT 205 205 G->E (IN DPY13(E225)).  
SQ SEQUENCE 302 AA; 30052 MW; 16BCSDAP14E0FE5 CRC64;

Query Match 12.5%; Score 169.5; DB 1; Length 302;  
Best local Similarity 33.5%; Pred. No. 0.00041;  
Matches 60; Conservative 7; Mismatches 53; Indels 59; Gaps 11;

DY 2 ADNFSLHDLSGSG-----NPNQGMFGANGNDPAGAGGYPGA-STYGPQA- 48  
Db 79 AVRVRGRDDAAAGNGFNPDSGCRCGCCPGPQGAAPGAGK--PGRPGRXPAAGFPFGNGPKAP 136

DY 49 -----PPGAYP-QQAPPGAVHAGP-----AYPGAPAPGVYPP--PSGGAYP 89  
Db 137 QKPCEBITTPCKPCRCQGEPGA-PGLPDPDQKGKXGAGCGCPGSIDAPAEQGPKNVAP 195

DY 90 SSGQSPAPGAYA-----TGYYGAPAGPLIVYNLLPYGVVPR 127  
Db 196 --GXKPAAPAGPDGHDPACVEPVQKGPSTGTBPVG--PPGPSGPGANDGTPGQGPX 251

RESULT 48  
PEBI\_DROME STANDARD; PRT; 377 AA.  
AC Q9U6L5; Q9WOM1;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ejaculatory bulb specific protein I precursor (PEB-me).  
GN PEB OR BCDNA:GH06048 OR CG268.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RS SEQUENCE FROM N.A.  
STRAIN=Berkely;  
MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Gehlker S.E., Holt R.A., Evans C.A., Goodyne J.D.,  
AMA Amaratilake P.G., Scherer S.F., Tsi R.P.W., Hoehnle B.J., Gatti D.



RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Tardell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Bantu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.V., Berno P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Flier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemtson J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Klip D., Lai Z.,  
 RA Lako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Lian X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RA Science 287:2185-2195(2000).  
 [2]  
 RP REVISIONS.  
 RA MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bertemont B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Sny S.O.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RA systematic review.";  
 RA Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Head;  
 RA MEDLINE=20196012; PubMed=10731138;  
 RA Rubin G.M., Hong L., Brokstein P., Evans-Hoim M., Friese E.,  
 RA Stapleton M., Harvey D.A.;  
 RA "A *Drosophila* complementary DNA resource.";  
 RA Science 287:2222-2224(2000).  
 [4]  
 RP SEQUENCE OF 77-90 AND 101-109, FUNCTION, SUBCELLULAR LOCATION, TISSUE  
 RP SPECIFICITY, AND VARIANT.  
 RC STRAIN=Canton-S;  
 RA MEDLINE=21167515; PubMed=11267893;  
 RA Lung O., Wolther M.P.;  
 RA "Identification and characterization of the major *Drosophila*  
 RA melanogaster mating plug protein.";  
 RA Insect Biochem. Mol. Biol. 31:543-551(2001).  
 CC -1- FUNCTION: Major protein component of the posterior mating plug.  
 CC Accessory gland proteins constitute, or are required for formation  
 CC of the anterior mating plug. Posterior mating plug forms before  
 CC sperm transfer and the anterior mating plug is formed after the  
 CC start of mating.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Specifically expressed in the ejaculatory bulb

CC and seminal fluid. Detected in mated females 3 minutes after the  
 CC start of mating, and for at least 3 hours after the start of  
 CC mating.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: A6003466; AAF47319.2; -;  
 DR EMBL: AF184225; AAD5736.1; -;  
 DR FlyBase, FBgn0004181; Peb.  
 KW Behavior; Signal; Polymorphism.  
 FT SIGNAL 1 20  
 FT CHAIN 21 377 POTENTIAL EJACULATORY BULB SPECIFIC PROTEIN I.  
 FT DOMAIN 116 256 PRO-RICH.  
 FT DOMAIN 159 377 GLY-RICH.  
 FT VARIANT 84  
 SO SEQUENCE 377 AA; 37819 MW; D020F027BD96B221 CRC64;  
 Query Match 12.5%; Score 169.5; DB 1; Length 377;  
 Best Local Similarity 34.9%; Pred. No. 0.0005;  
 Matches 53; Conservative 10; Mismatches 46; Indels 43; Gaps 10;  
 QY 19 PQGNP-----GAMGNP-----AGAGVPGASYPGPGAPGAPGAPGAY 62  
 DB 129 PLGMLRPLPGPLVPMRPMPLINSPIRPGPLPGGPPSGPSPGSPS- 184  
 QY 63 HGAGVAPGAPAP-GVPPGPPSGGAYPSGQSPAPAVATG---PYGAPAPLIVPNTL 118  
 DB 185 ---GSPGSGSPGSPGSPGSGS-PGG-PSPGSPGPGPFGGSPSGGPGPQFPWIL 238  
 QY 119 -----PLPGVVRMLITLITLGYKRNA 140  
 DB 239 GGPFRNRPGRPPGILPGHLD--GSVVPNS 267  
 RESULT 49  
 ANXB HUMAN STANDARD; PRT; 505 AA.  
 ID ANXB\_HUMAN  
 AC P50935;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Annexin A11 (Annexin XI) (calyculin-associated annexin 50) (CAP-50)  
 DE (56 kDa autoantigen).  
 GN ANXA11 OR ANX11.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Teratocarcinoma;  
 RA MEDLINE=94140847; PubMed=7508441;  
 RA Wiskai Y., Pruijn G.J.M., van der Kemp A.W., van Venrooi W.J.;  
 RA "The 56k autoantigen is identical to human annexin XI.";  
 RA J. Biol. Chem. 269:4240-4246(1994).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20469408; PubMed=11013079;  
 RA Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,  
 RA Fernandez M.P.;  
 RA "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous  
 RA annexins and source of orthologous cDNA isoforms.";  
 RA Genomics 69:95-103(2000).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;  
 RA MEDLINE=22388257; PubMed=12477932;



RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bock S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Holik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Heaton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.B.,  
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND POSSIBLY NUCLEAR.  
 CC -1- DOMAIN: A pair of annexin repeats may form one binding site for  
 CC calcium and phospholipid.  
 CC -1- DISBASE: Antibodies against ANXA11 are present in sera from  
 CC patients with various autoimmune diseases, predominantly in sera  
 CC from patients with rheumatoid arthritis, systemic lupus  
 CC erythematosus, or Sjogren's syndrome.  
 CC -1- SIMILARITY: Belongs to the annexin family.  
 CC -1- SIMILARITY: Contains 4 annexin repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL; L19605; AAA19734.1; -;  
 DR EMBL; AJ278463; CAB94996.1; -;  
 DR EMBL; AJ278464; CAB94996.1; -;  
 DR EMBL; AJ278465; CAB94997.1; -;  
 DR EMBL; BC007564; AAH07564.1; -;  
 DR PIR; A53152; A53152.  
 DR HSSP; P13214; IACW.  
 DR Genew; HGNC:535; ANXA11.  
 DR MIM; 602572; -;  
 DR GO; GO:0005737; C:cytoplasm; TAS.  
 DR GO; GO:0005635; C:nuclear membrane; NAS.  
 DR GO; GO:0005654; C:nucleoplasm; NAS.  
 DR GO; GO:0005543; F:phospholipid binding; TAS.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0006955; P:immune response; TAS.  
 DR InterPro; IPR001464; Annexin.  
 DR Pfam; PF00191; annexin; 4.  
 DR PRINTS; PR00196; ANNEXIN.  
 DR ProDom; PD000143; Annexin; 4.  
 DR SMART; SM00335; ANX; 4.  
 DR PROSITE; PS00223; ANNEXIN; 4.  
 KM Annexin; Calcium/phospholipid-binding; Repeat; Polymorphism.  
 FT REPEAT 209 269 ANNEXIN 1.  
 FT REPEAT 281 341 ANNEXIN 2.  
 FT REPEAT 365 425 ANNEXIN 3.  
 FT REPEAT 440 500 ANNEXIN 4.  
 FT VARIANT 230 230 R -> C (in dbSNP:1049550).  
 FT VARIANT 457 457 /FTID=VAR\_012006.  
 FT VARIANT 457 457 I -> V (in dbSNP:1802932).  
 FT /FTID=VAR\_012007.  
 SQ SEQUENCE 505 AA; 54389 MW; 4ADCA08F270BEE4 CRC64;

QY 11 LSGSGNPNPGQWPGAWGNQDAGAGVP--GASYPGYGGAPP-----AYPGQ----- 56  
 Db 1 MSVPGYPPPG-----GYPPAPAGGPGWGAAYPPPSMPPRIGLDNVAATAGQNDYLS 55  
 QY 57 ---APGAYHGA--PGATGAPAPGVYPPGPPSPGAYPSSGQPSAAGATATGYGAPAG 110  
 Db 56 GMAANSGTGGGAMPMLYGPAGPAGYPPVPPGGFGQPPSAQGFVPP-----YGMVPPPG 111  
 QY 111 --PLIVRYNLPPLPGGVPRMLITLGTVKDNA 140  
 Db 112 NPESRMSPYPPYGPAPVPQGMPPPGQPPGA 143  
 RESULT 50  
 ID CA14\_HUMAN STANDARD; PRT; 1669 AA.  
 AC P02462;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, last sequence update)  
 DT 10-OCT-2003 (Rel. 42, last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 OS COL4A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89340433; PubMed=2701944;  
 RA Souminen R., Hucartari M., Ganguly A., Prockop D.J., Tryggvason K.;  
 RT "Structural organization of the gene for the alpha 1 chain of human  
 RT type IV collagen.";  
 RL J. Biol. Chem. 264:13565-13571(1989).  
 RN [2]  
 RP SEQUENCE OF 46-1257 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88083584; PubMed=3691802;  
 RA Souminen R., Haka-Risku T., Prockop D.J., Tryggvason K.;  
 RT "Complete primary structure of the alpha 1-chain of human basement  
 RT membrane (type IV) collagen.";  
 RL FEBS Lett. 225:188-194(1987).  
 RN [3]  
 RP SEQUENCE OF 1-943 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88029471; PubMed=3311751;  
 RA Brazel D., Oberhauser I., Dieringer H., Babel W., Glanville R.W.,  
 RA Deitzmann R., Kuehn K.;  
 RT "Completion of the amino acid sequence of the alpha 1 chain of human  
 RT basement membrane collagen (type IV) reveals 21 non-triplet  
 RT interruptions located within the collagenous domain.";  
 RL Eur. J. Biochem. 168:529-536(1987).  
 RN [4]  
 RP SEQUENCE OF 28-243.  
 RX MEDLINE=86004708; PubMed=4043082;  
 RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;  
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking  
 RT region (7S domain) of the alpha 1 (IV) chain of human basement  
 RT membrane collagen.";  
 RL Eur. J. Biochem. 152:213-219(1985).  
 RN [5]  
 RP SEQUENCE OF 534-1447.  
 RX MEDLINE=85003629; PubMed=6434307;  
 RA Babel W., Glanville R.W.;  
 RT "Structure of human-basement-membrane (type IV) collagen. Complete  
 RT amino-acid sequence of a 914-residue-long pepsin fragment from the  
 RT alpha 1(IV) chain.";  
 RL Eur. J. Biochem. 143:545-556(1984).  
 RN [6]  
 RP SEQUENCE OF 1256-1669 FROM N.A.  
 RX MEDLINE=85207819; PubMed=2581699;  
 RA Philanjaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;

RT "cDNA clones coding for the pro-alpha(IV) chain of human type IV  
RT procollagen reveal an unusual homology of amino acid sequences in two  
RT halves of the carboxyl-terminal domain."  
RT J. Biol. Chem. 260:7681-7687(1985).  
RN [7]  
RP SEQUENCE OF 1259-1669 FROM N.A.  
RA MEDLINE=85216555; PubMed=2582422;  
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
RA Kefalides N.A., Myers J.C.;  
RT "Restricted homology between human alpha 1 type IV and other  
RT procollagen chains."  
RT Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).  
RN [8]  
RP SEQUENCE OF 1-28 FROM N.A.  
RA MEDLINE=89034231; PubMed=3182844;  
RA Soiminen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
RT collagen are divergently encoded on opposite DNA strands and have an  
RT overlapping promoter region."  
RT J. Biol. Chem. 263:17217-17220(1988).  
RN [9]  
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
RC TISSUE=Placenta;  
RA MEDLINE=8905112; PubMed=2844531;  
RA Siebold B., Deutzmann R., Kuehn K.;  
RT "The arrangement of intra- and intermolecular disulfide bonds in the  
RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
RT of basement-membrane type IV collagen."  
RT Eur. J. Biochem. 176:617-624(1988).  
RL -I- FUNCTION: Type IV collagen is the major structural component of  
CC glomerular basement membranes (GBM), forming a 'chicken-wire'  
CC meshwork together with laminins, proteoglycans and entactin/  
CC nidogen.  
CC -I- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-  
CC alpha 6(IV), each of which can form a triple helix structure  
CC with 2 other chains to generate type IV collagen network.  
CC -I- DOMAIN: Alpha chains of type IV collagen have a noncollagenous  
CC domain (NC1) at their C-terminus, frequent interruptions of the  
CC G-X-Y repeats in the long central triple-helical domain (which may  
CC cause flexibility in the triple helix), and a short N-terminal  
CC triple-helical 7S domain.  
CC -I- PTM: Lysines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
CC -I- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -I- PTM: Type IV collagens contain numerous cysteine residues which  
CC are involved in inter- and intramolecular disulfide bonding. 12 of  
CC these, located in the NC1 domain, are conserved in all known type  
CC IV collagens.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL, M26576; AAA53098.1; JOINED.  
DR EMBL, J04217; AAA53098.1; JOINED.  
DR EMBL, M26550; AAA53098.1; JOINED.  
DR EMBL, M26540; AAA53098.1; JOINED.  
DR EMBL, M26542; AAA53098.1; JOINED.  
DR EMBL, M26543; AAA53098.1; JOINED.  
DR EMBL, M26544; AAA53098.1; JOINED.  
DR EMBL, M26545; AAA53098.1; JOINED.  
DR EMBL, M26546; AAA53098.1; JOINED.  
DR EMBL, M26547; AAA53098.1; JOINED.  
DR EMBL, M26537; AAA53098.1; JOINED.  
DR EMBL, M26538; AAA53098.1; JOINED.  
DR EMBL, M26548; AAA53098.1; JOINED.  
DR EMBL, M26549; AAA53098.1; JOINED.  
DR EMBL, M26551; AAA53098.1; JOINED.

DR EMBL, M26552; AAA53098.1; JOINED.  
DR EMBL, M26553; AAA53098.1; JOINED.  
DR EMBL, M26554; AAA53098.1; JOINED.  
DR EMBL, M26555; AAA53098.1; JOINED.  
DR EMBL, M26556; AAA53098.1; JOINED.  
DR EMBL, M26557; AAA53098.1; JOINED.  
DR EMBL, M26558; AAA53098.1; JOINED.  
DR EMBL, M26559; AAA53098.1; JOINED.  
DR EMBL, M26560; AAA53098.1; JOINED.  
DR EMBL, M26561; AAA53098.1; JOINED.  
DR EMBL, M26562; AAA53098.1; JOINED.  
DR EMBL, M26563; AAA53098.1; JOINED.  
DR EMBL, M26564; AAA53098.1; JOINED.  
DR EMBL, M26565; AAA53098.1; JOINED.  
DR EMBL, M26566; AAA53098.1; JOINED.  
DR EMBL, M26567; AAA53098.1; JOINED.  
DR EMBL, M26568; AAA53098.1; JOINED.  
DR EMBL, M26569; AAA53098.1; JOINED.  
DR EMBL, M26570; AAA53098.1; JOINED.  
DR EMBL, M26571; AAA53098.1; JOINED.  
DR EMBL, M26572; AAA53098.1; JOINED.  
DR EMBL, M26573; AAA53098.1; JOINED.  
DR EMBL, M26574; AAA53098.1; JOINED.  
DR EMBL, M26575; AAA53098.1; JOINED.  
DR EMBL, Y00706; CAA68698.1; -.  
DR EMBL, X05561; CAA29075.1; -.  
DR EMBL, M10940; AAA52006.1; -.  
DR EMBL, M11315; AAA52042.1; -.  
DR PIR, S16876; CGH04B.  
DR Genew; HGNC:2202; COL4A1.  
DR MIM, 120130; -.  
DR InterPro; IPR008161; C1g helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagen4\_C.  
DR Pfam; PF01431; C4; 2.  
DR Pfam; PF01391; Collagen\_24.  
DR Pfam; PF000007; C1g helix; 6.  
DR ProDom; PD0003923; ProcollagenC4; 1.  
DR SMART; SM00111; C4; 2.  
DR Extracellular matrix; Connective tissue; Basement membrane;  
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 27  
FT PROPEP 28 172  
FT CHAIN 173 1669  
FT DOMAIN 173 1440  
FT DOMAIN 1441 1669  
FT CARBOHYD 126 126  
FT DISULFID 1460 1551  
FT DISULFID 1493 1548  
FT DISULFID 1505 1511  
FT DISULFID 1570 1665  
FT DISULFID 1604 1662  
FT DISULFID 1616 1662  
FT DISULFID 237 238  
FT CONFLICT 241 241  
FT CONFLICT 319 319  
FT CONFLICT 719 719  
FT CONFLICT 837 837  
FT CONFLICT 842 842  
FT CONFLICT 896 896  
FT CONFLICT 904 904  
FT CONFLICT 914 914  
FT CONFLICT 998 998  
FT CONFLICT 1010 1010  
FT CONFLICT 1012 1012  
FT CONFLICT 1358 1358  
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFB8B8A84 CRC64;  
Query Match 12.5%; Score 169; DB 1; Length 1669;  
Best Local Similarity 37.7%; Pred. No. 0.0021;

	Matches	55; Conservative	9; Mismatches	38; Indels	44; Gaps	11;
QY	14	SGNPNPQGMPGAWGNQPAAGAGYPGASY-----	PGYPGQAPPAYPGQAPPAY	62		
Db	509	AGVPPGQCTPLIG-QP-GAKGEPPGFYFDLRLKGDKDDPGFPQG--PG-MGRRA-----		558		
QY	63	HGAPG--AYGAPAPGVYP-----GPPSG--PGAYPSSGQPSAPGAYATGPY---		105		
Db	559	-GSPGRDGHPIPGPKGSPGSAVGLKGERGPPGVGFPGRDPTGPPDPGYGPAGPIGDK		617		
QY	106	---GAPAGPLT--VPYNLPLPGGVYP		126		
Db	618	GQAGFPGGGSPGLPGPKGEPKIVP		643		

Search completed: August 23, 2004, 14:37:32  
 Job time : 15 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: August 23, 2004, 14:33:32 ; Search time 22 Seconds  
(without alignments)  
1075.596 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357

Sequence: 1 MADNPSLHDAISGSGNPNPQ.....ISKLGISGIDILTSASYTMI 246

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 65 summaries

Database :

PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1315	96.9	250	2	A35820 galactin 3 - human
2	1082	79.7	264	2	A28651 galactose-specific
3	1082	79.7	264	2	A45983 lactose-binding le
4	1076.5	79.3	294	2	A49688 lactose-binding le
5	1075	79.2	262	2	A54889 IGF-binding protei
6	1049.5	77.3	245	2	A54909 carboxylate-bind
7	1024	75.5	242	2	JC4300 galactin-3 - rabbi
8	437	32.2	139	2	S08576 lectin - mouse (fr
9	254	18.7	323	2	A55664 lectin L-36 - pig
10	250.5	18.5	145	2	A55932 galactin-5 - rat
11	236.5	17.4	324	2	A46631 lactose-binding le
12	223.5	16.5	317	2	JC6147 prostate carcinoma
13	218.5	16.1	316	2	A55975 galactin-8 - rat
14	211	15.5	136	2	I55469 galactin-7 - human
15	193	14.2	1049	1	CG8078 collagen alpha 1(I
16	192.5	14.0	463	2	S29170 annexin VII - mous
17	190.5	14.0	177	2	S65780 glycine/proline-ri
18	189.5	14.0	329	2	I32783 hypothetical prote
19	188.5	13.9	279	2	I37216 beta-galactoside-b
20	188.5	13.9	285	2	T26325 hypothetical prote
21	188	13.9	744	2	S15435 collagen alpha 1(I
22	187.5	13.8	313	2	T22828 hypothetical prote
23	186.5	13.7	179	2	A85217 hypothetical prote
24	186.5	13.7	277	2	T04441 amnextrin VII, long
25	186	13.7	488	1	LUMU7 collagen alpha 1(I
26	185.5	13.6	488	2	A27353 hypothetical prote
27	184	13.6	743	1	T19564 collagen alpha 1(I
28	183.5	13.5	743	1	S23779 collagen alpha 1(I
29	183	13.5	299	2	T25407 hypothetical prote

## ALIGNMENTS

30	182.5	13.4	317	2	T29960 hypothetical prote
31	181.5	13.4	304	2	T26185 hypothetical prote
32	181.5	13.4	304	2	T26185 hypothetical prote
33	180.5	13.3	356	2	T22827 hypothetical prote
34	180	13.3	744	1	A34246 collagen alpha 1(I
35	180	13.3	744	1	S23298 collagen alpha 1(I
36	179	13.2	316	2	S08169 collagen col-12 pr
37	179	13.2	316	2	S08170 collagen col-13 pr
38	178	13.1	1262	2	T22523 hypothetical prote
39	177	13.0	299	2	T24833 hypothetical prote
40	176.5	13.0	304	2	T22482 hypothetical prote
41	175.5	12.9	1357	2	T29265 hypothetical prote
42	175	12.9	569	2	S42886 collagen - silkwor
43	174.5	12.9	290	2	T24586 hypothetical prote
44	174.5	12.9	296	2	A31219 collagen 1 - Caeno
45	174.5	12.9	299	2	T29956 hypothetical prote
46	174	12.8	285	2	R88281 protein ZK892.1 (I
47	174	12.8	285	2	T29982 hypothetical prote
48	173.5	12.8	266	2	T22706 hypothetical prote
49	173.5	12.8	299	2	T22705 hypothetical prote
50	173.5	12.8	512	2	S70644 amnextrin VII - Afr1
51	173.5	12.8	754	2	A55267 collagen alpha 5(I
52	172.5	12.7	1464	2	S59856 collagen alpha 1(I
53	172	12.7	283	2	T29980 hypothetical prote
54	172	12.7	301	2	T21314 hypothetical prote
55	172	12.7	301	2	B31219 collagen 2 - Caeno
56	172	12.7	675	2	S20819 collagen alpha 3(I
57	171.5	12.6	464	2	A47655 splicosome-associ
58	171.5	12.6	1712	1	CGH028 collagen alpha 2(I
59	171	12.6	290	2	T24590 hypothetical prote
60	171	12.6	416	2	T32458 hypothetical prote
61	170.5	12.6	314	2	T32985 hypothetical prote
62	170	12.5	305	2	T20906 collagen alpha 2(I
63	170	12.5	469	2	A24450 collagen alpha 2(I
64	170	12.5	1466	1	CGH07L collagen alpha 1(I
65	169.5	12.5	302	2	A31921 collagen dpy-13 pr

RESULT 1  
A35820  
galactin 3 - human  
N:Alternative names: CBP 35; epithelial-specific lectin 35; galactoside-binding lectin L-3  
A:Species: Homo sapiens (man)  
C:Date: 12-Oct-1990 #sequence revision 12-Oct-1990 #text\_change 20-Apr-2000  
R:Accession: A35820; J00916; A47473; A36071; A45800  
R:Robertson, M.W.; Albrandt, K.; Keller, D.; Liu, F.T.  
Biochemistry 29, 8093-8100, 1990  
A:Title: Human IGF-binding protein: a soluble lectin exhibiting a highly conserved inte  
A:Reference number: A35820; MUID:91084480; PMID:2261464  
A:Accession: A35820  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-250 <ROB>  
R:Oda, Y.; Leffler, H.; Sakakura, Y.; Kasai, K.; Baronides, S.H.  
Gene 99, 279-283, 1991  
A:Title: Human breast carcinoma cDNA encoding a galactoside-binding lectin homologous t  
A:Reference number: J00916; MUID:91216471; PMID:2022338  
A:Accession: J00916  
A:Molecule type: mRNA  
A:Residues: 1-250 <ODA>  
A:Cross-references: GB:M56682; NID:g186921; PIDN:AAA36163.1; PID:g186922  
R:Lotz, M.M.; Andrews Jr., C.W.; Korzeilus, C.A.; Lee, E.C.; Steele Jr., G.D.; Clarke,  
Proc. Natl. Acad. Sci. U.S.A. 90, 3466-3470, 1993  
A:Title: Decreased expression of Mac-2 (carbohydrate binding protein 35) and loss of it  
A:Reference number: A47473; MUID:93234518; PMID:7682704  
A:Accession: A47473  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-63, 'P', 65-97, 'T', 99-250 <LOT>

A:Cross-references: GB:S59012; NID:9299602; PTDN:AA626229.1; PID:9299602  
 A:Experimental source: normal colonic mucosa, colon carcinoma, cell line clone A  
 A>Note: sequence extracted from NCBI backbone (NCBIN:129689, NCBIPI:129692)  
 R:Cherayil, B.J.; Chaitovitz, S.; Monq, C.; Pillai, S.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 7324-7328, 1990  
 A>Title: Molecular cloning of a human macrophage lectin specific for galactose.  
 A:Reference number: A36071; MUID:90384999; PMID:2402511  
 A:Accession: A36071  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-32, 'Q', '34', 'L', '37', 'RGLSMGL', '46', 'RAGT', '51', 'R', '53-63', 'P', '65-87', '89-250 <CHE>  
 A:Cross-references: GB:M35368; NID:g1196441  
 A>Note: The sequence is revised in GenBank entry HUMMAC2, release 113.0. PTDN:AAA8086.1.1  
 R:Raz, A.; Carmi, P.; Raz, T.; Hogan, V.; Mohamed, A.; Wolman, S.R.  
 Cancer Res. 51, 2173-2178, 1991  
 A>Title: Molecular cloning and chromosomal mapping of a human galactoside-binding protease  
 A:Reference number: A49800; MUID:91183475; PMID:2009535  
 A:Accession: A49800  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-32, 'Q', '34', 'LPGASYPGAPAGTPTGHSWTAPGATMEHVELIRSTCTWSLRQ', '86-104', 'A', '106', 'W', '108-110', 'L', '112-113', 'L', '115-116', 'L', '118-119', 'L', '121-122', 'L', '124-125', 'L', '127-128', 'L', '130-131', 'L', '133-134', 'L', '136-137', 'L', '139-140', 'L', '142-143', 'L', '145-146', 'L', '148-149', 'L', '151-152', 'L', '154-155', 'L', '157-158', 'L', '160-161', 'L', '163-164', 'L', '166-167', 'L', '169-170', 'L', '172-173', 'L', '175-176', 'L', '178-179', 'L', '181-182', 'L', '184-185', 'L', '187-188', 'L', '190-191', 'L', '193-194', 'L', '196-197', 'L', '199-200', 'L', '202-203', 'L', '205-206', 'L', '208-209', 'L', '211-212', 'L', '214-215', 'L', '217-218', 'L', '220-221', 'L', '223-224', 'L', '226-227', 'L', '229-230', 'L', '232-233', 'L', '235-236', 'L', '238-239', 'L', '241-242', 'L', '244-245', 'L', '247-248', 'L', '250-251', 'L', '253-254', 'L', '256-257', 'L', '259-260', 'L', '262-263', 'L', '265-266', 'L', '268-269', 'L', '271-272', 'L', '274-275', 'L', '277-278', 'L', '280-281', 'L', '283-284', 'L', '286-287', 'L', '289-290', 'L', '292-293', 'L', '295-296', 'L', '298-299', 'L', '301-302', 'L', '304-305', 'L', '307-308', 'L', '310-311', 'L', '313-314', 'L', '316-317', 'L', '319-320', 'L', '322-323', 'L', '325-326', 'L', '328-329', 'L', '331-332', 'L', '334-335', 'L', '337-338', 'L', '340-341', 'L', '343-344', 'L', '346-347', 'L', '349-350', 'L', '352-353', 'L', '355-356', 'L', '358-359', 'L', '361-362', 'L', '364-365', 'L', '367-368', 'L', '370-371', 'L', '373-374', 'L', '376-377', 'L', '379-380', 'L', '382-383', 'L', '385-386', 'L', '388-389', 'L', '391-392', 'L', '394-395', 'L', '397-398', 'L', '400-401', 'L', '403-404', 'L', '406-407', 'L', '409-410', 'L', '412-413', 'L', '415-416', 'L', '418-419', 'L', '421-422', 'L', '424-425', 'L', '427-428', 'L', '430-431', 'L', '433-434', 'L', '436-437', 'L', '439-440', 'L', '442-443', 'L', '445-446', 'L', '448-449', 'L', '451-452', 'L', '454-455', 'L', '457-458', 'L', '460-461', 'L', '463-464', 'L', '466-467', 'L', '469-470', 'L', '472-473', 'L', '475-476', 'L', '478-479', 'L', '481-482', 'L', '484-485', 'L', '487-488', 'L', '490-491', 'L', '493-494', 'L', '496-497', 'L', '499-500', 'L', '502-503', 'L', '505-506', 'L', '508-509', 'L', '511-512', 'L', '514-515', 'L', '517-518', 'L', '520-521', 'L', '523-524', 'L', '526-527', 'L', '529-530', 'L', '532-533', 'L', '535-536', 'L', '538-539', 'L', '541-542', 'L', '544-545', 'L', '547-548', 'L', '550-551', 'L', '553-554', 'L', '556-557', 'L', '559-560', 'L', '562-563', 'L', '565-566', 'L', '568-569', 'L', '571-572', 'L', '574-575', 'L', '577-578', 'L', '580-581', 'L', '583-584', 'L', '586-587', 'L', '589-590', 'L', '592-593', 'L', '595-596', 'L', '598-599', 'L', '601-602', 'L', '604-605', 'L', '607-608', 'L', '610-611', 'L', '613-614', 'L', '616-617', 'L', '619-620', 'L', '622-623', 'L', '625-626', 'L', '628-629', 'L', '631-632', 'L', '634-635', 'L', '637-638', 'L', '640-641', 'L', '643-644', 'L', '646-647', 'L', '649-650', 'L', '652-653', 'L', '655-656', 'L', '658-659', 'L', '661-662', 'L', '664-665', 'L', '667-668', 'L', '670-671', 'L', '673-674', 'L', '676-677', 'L', '679-680', 'L', '682-683', 'L', '685-686', 'L', '688-689', 'L', '691-692', 'L', '694-695', 'L', '697-698', 'L', '700-701', 'L', '703-704', 'L', '706-707', 'L', '709-710', 'L', '712-713', 'L', '715-716', 'L', '718-719', 'L', '721-722', 'L', '724-725', 'L', '727-728', 'L', '730-731', 'L', '733-734', 'L', '736-737', 'L', '739-740', 'L', '742-743', 'L', '745-746', 'L', '748-749', 'L', '751-752', 'L', '754-755', 'L', '757-758', 'L', '760-761', 'L', '763-764', 'L', '766-767', 'L', '769-770', 'L', '772-773', 'L', '775-776', 'L', '778-779', 'L', '781-782', 'L', '784-785', 'L', '787-788', 'L', '790-791', 'L', '793-794', 'L', '796-797', 'L', '799-800', 'L', '802-803', 'L', '805-806', 'L', '808-809', 'L', '811-812', 'L', '814-815', 'L', '817-818', 'L', '820-821', 'L', '823-824', 'L', '826-827', 'L', '829-830', 'L', '832-833', 'L', '835-836', 'L', '838-839', 'L', '841-842', 'L', '844-845', 'L', '847-848', 'L', '850-851', 'L', '853-854', 'L', '856-857', 'L', '859-860', 'L', '862-863', 'L', '865-866', 'L', '868-869', 'L', '871-872', 'L', '874-875', 'L', '877-878', 'L', '880-881', 'L', '883-884', 'L', '886-887', 'L', '889-890', 'L', '892-893', 'L', '895-896', 'L', '898-899', 'L', '901-902', 'L', '904-905', 'L', '907-908', 'L', '910-911', 'L', '913-914', 'L', '916-917', 'L', '919-920', 'L', '922-923', 'L', '925-926', 'L', '928-929', 'L', '931-932', 'L', '934-935', 'L', '937-938', 'L', '940-941', 'L', '943-944', 'L', '946-947', 'L', '949-950', 'L', '952-953', 'L', '955-956', 'L', '958-959', 'L', '961-962', 'L', '964-965', 'L', '967-968', 'L', '970

A:Title: Chem 263, 6009-6011, 1988  
 A:Title: Carbohydrate binding protein 35. Complementary DNA sequence reveals homology with  
 A:Reference number: A28651; MUID:88198129; PMID:3360772  
 A:Accession: A28651  
 A:Molecule type: mRNA  
 A:Residues: 'R', 3-264 <JTA>  
 A:Cross-references: EMBL:J03723  
 R:Raz, A.; Pazerini, G.; Carni, P.  
 Cancer Res. 49, 3489-3493, 1989  
 A:Title: Identification of the metastasis-associated, galactoside-binding lectin as a ch  
 A:Reference number: A57385; MUID:89275058; PMID:2525069  
 A:Accession: A57385  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Cross-references: GB:J16074; NID:952850; PIDN:CAA34206.1; PID:952851  
 A:Note: authors translated the codon GAA for residue 219 as Ala, GAC for residue 221 as C  
 J. Biol. Chem. 265, 7097-7099, 1990  
 R:Woo, H.U.; Shaw, L.M.; Messier, J.M.; Mecurio, A.M.  
 A:Title: The major non-integrin laminin binding protein of macrophages is identical to c  
 A:Reference number: A55185; MUID:90236991; PMID:2332426  
 A:Accession: A55185  
 A:Molecule type: protein  
 A:Residues: 159-163, 166-175, 214-226 <MOO>  
 C:Superfamily: beta-galactoside-binding lectin  
 C:Keywords: lectin; phosphoprotein  
 Query Match 79.7%; Score 1082; DB 2; Length 264;  
 Best Local Similarity 78.0%; Pred. No. 9e-69;  
 Matches 209; Conservative 16; Mismatches 17; Indels 26; Gaps 8;  
 QY 1 MANNPSLHDALISGSNPDPQGMGAMGNOPAGAGYGPASVYG-YYGQAPPGAYPGQAPP 59  
 1 MADPSLINDALAGSNPNPDGPGAMGNOP-GAGGYPGAYPGAYPGQAPPGAYPGAPP 59  
 QY 60 GAYHG-----AAGAYPGAPAPGAYPGPPSGPGAYPSGQPSAGAY----- 100  
 DB 60 GAYPGQAPPSAYPGPTAGAYPGPTAPGAYPGQAP-PGAF-P-GDPGAPGAYPGQSGGYP 116  
 QY 101 ATGEGYAGAPAGLIVEYDLPPLPGGVPRMLITLITGKPNANLIALDPQRGNDAFHP-PR 159  
 DB 117 AAGGYGAPAGLIVEYDLPPLPGGVPRMLITIMGTVPKANRILVDFRRGNDAFHPNPR 176  
 QY 160 FNNNNRRIVYNTKLDNNWGSEEROSVFPESGKPKLOVYEPDHFXYAVNDNAH-LQYN 218  
 DB 177 FNNNNRRIVYNTKLDNNWGSEEROSVFPESGKPKLOVLEADHFKAANDAHLLQYN 236  
 QY 219 HRVKKNEISKLGISGDIDLTASATMI 246  
 DB 237 HRMKNLREISQLGISGDITLTSANMI 264  
 RESULT 3  
 A45983  
 C:Accession: A45983  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #ext\_change 03-May-1996  
 R:Rosenberg, I.M.; Iyer, R.; Cherayil, B.; Chiodino, C.; Pillai, S.  
 U. Biol. Chem. 268, 12393-12400, 1993  
 A:Title: Structure of the murine Mac-2 gene. Splice variants encode proteins lacking func  
 A:Reference number: A45983; MUID:93286070; PMID:8509379  
 A:Accession: A45983  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-264 <ROS>  
 A:Cross-references: GB:L08649  
 C:Genetics:  
 C:Superfamily: beta-galactoside-binding lectin  
 Query Match 79.7%; Score 1082; DB 2; Length 264;  
 Best Local Similarity 78.6%; Pred. No. 9e-69;





A:Reference: S07162; MUID:88080093; PMID:3335026  
A:Accession: S08576  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-139 <RAZ>  
C:Superfamily: beta-galactoside-binding lectin

Query Match      32.2%; Score 437; DB 2; Length 139;  
Best Local Similarity 75.9%; Pred. No. 5, 6e-24;  
Matches 85; Conservative 6; Mismatches 19; Indels 2; Gaps 2;

OY 104 PYGAPAGPLIVPYNLPDGGVPRMLITLIGTVKPNANRIALDFQGNDAFHF-PPNFE 162  
DB 22 PLMCRWTLTDVPDYDLPLPGGVMPRLITIMGTVKPNANIIVLDFRRGNDVAFFHPPRNE 81  
OY 163 NNRRIVCNTXLDNNMGREEROSVFPPESGKPFKIQLVEPDHPKAVVNDAH 214  
DB 82 NNRRIVCNTKODNNMGKEEROSAFPEECGNPFKIQLVLAADHSGCG-HDAH 132

RESULT 9  
A:Accession: A55664  
lectin L-36 - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #ext\_change 29-Sep-1999  
C:Accession: A55664  
J:Chiu, M.L.; Parry, D.A.D.; Feldman, S.R.; Klapper, D.G.; O'Keefe, E.J.  
J: Biol. Chem. 269, 31770-31776, 1994  
A>Title: An adherens junction protein is a member of the family of lactose-binding lectins  
A:Reference number: A55664; MUID:95081129; PMID:7989350  
A:Accession: A55664  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-323 <CHI>  
C:Cross-references: Gen: X79303; NID:9623345; PIDN:CAA55884.1; PUD:g623346  
C:Superfamily: lactose-binding lectin L-36

Query Match      18.7%; Score 254; DB 2; Length 323;  
Best Local Similarity 37.6%; Pred. No. 8, 6e-11;  
Matches 68; Conservative 22; Mismatches 77; Indels 14; Gaps 8;

OY 71 GAPAPGYVPGSPGSGAYPSGQPAPGAYATGPAPGLIIVPYNLPDPGVMPRLI 130  
DB 152 GQAPAS--PGEMPNG-YPGEGKNHQPCNLFCMEGAPTFNPVPYPKTRLGGSLVARRTI 208  
OY 131 TILGTAKPNANRIALDFQRGN-DVAEHP-PFNENNRRIVVCNTKLDDNWGEERQSVP 187  
DB 209 VIKSYVPSGKSIVINFKVGSGDVALLHNPRLTG--IVRNSTYLNGKGAEERKSSF 265  
OY 188 -PFGSGKPFKIQLVEPDHPKAVVNDAH-QYNHHVKKLNLSKIGSGIDLTSASYIM 245  
DB 266 NFPAAGQYFDLSIRKGLDRPKYVANGCHLFPFSHLISNFQGVDTLEIQGDVTU--SYVQ 322

OY 246 I 246  
DB 323 I 323

RESULT 10  
A:Accession: A55932  
galactin-5 - rat  
N:Alternate names: beta-galactoside binding lectin  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #ext\_change 21-Jan-2000  
C:Accession: A55932; PK0077  
J:Giltz, M.A.; Wieser, M.F.; Leffler, H.; Herrmann, J.; Xia, Y.R.; Massa, S.M.; Cooper, D.N.  
J: Biol. Chem. 270, 5032-5038, 1995  
A>Title: Sequence and mapping of galactin-5, a beta-galactoside-binding lectin, found in  
A:Reference number: A55932; MUID:95197487; PMID:7890611  
A:Accession: A55932  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-145 <GIT>

A:Cross-references: GB:L36862; NID:g727175; PIDN:AAQ42050.1; PID:g727176  
 R.Jung, S.K.; Fujimoto, D.  
 J. Biochem. 116: 547-553, 1994  
 A:Title: A novel beta-galactoside-binding lectin in adult rat kidney.  
 A:Reference number: PX0077; MUID:95155264; PMID:7852273  
 A:Accession: PX0077  
 A:Molecule type: protein  
 A:Residues: 7,'P',9-10,'T',12-19,'X',21-25;30-42;109-111,'N',113,'H',115,'VS',118-123,'X'  
 A:Experimental source: kidney  
 C:Comment: This protein exhibits activity to various saccharides and binds to Engelbreth  
 C:Genetics:  
 A:Gene: LGALS5  
 C:Superfamily: beta-galactoside-binding lectin  
 C:Keywords: acetylated amino end; lectin; monomer  
 E:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental

Query Match 18.5%; Score 250.5; DB 2; Length 145;  
 Best Local Similarity 40.4%; Pred. No. 6,4e-11;  
 Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;

QY 108 PAGPLIVPNLPLPGGVPRMLITLGTVKPNANRIALDFQGNVAFHF-PRENNNR 166  
 DB 9 PYPNLAIVPFTSINGLTPSKIVISGVVLSDAKFPQNLRCGDIAPHLPDEN--- 65  
 QY 167 VIVCNTKLDNNWGREERQ--SVFPESGKPKICQVLEPDHFKVAVNDAAH-LQYHRYVK 223  
 DB 66 AVVNTQQLNNSWGPERSLPGSMFPRGQRFVWLCEGCHKVAVDQGHICEYHRLMN 125  
 QY 224 LNEISKLGISGDIIDLT 239  
 DB 126 LFDINTLEVAGDIQLT 141

RESULT 11  
 A4631  
 Lactose-binding lectin L-36 - rat  
 N:Alternate names: galectin-4  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 29-Sep-1999  
 C:Accession: A4631; S69096  
 R:Oda, Y.; Herrmann, U.; Gilt, M.A.; Turck, C.W.; Burlingame, A.L.; Barondes, S.H.; Left  
 J. Biol. Chem. 268, 5929-5939, 1993  
 A:Title: Soluble lactose-binding lectin from rat intestine with two different carbohydrate  
 A:Reference number: A4631; MUID:93194902; PMID:8449956  
 A:Accession: A4631  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-324 <ODA>  
 A:Cross-references: GB:M73553; NID:g294571; PIDN:AAA41505.1; PID:g294572  
 A:Experimental source: intestine  
 A:Note: sequence extracted from NCBI backbone (NCBIN:128409, NCBIPI:128410)  
 R:Jardry, F.; Devillier, P.; Louisot, P.; Martin, A.  
 FEBS Lett. 359, 169-172, 1995  
 A:Title: Purification and characterization of the N-terminal domain of galectin-4 from r  
 A:Reference number: S69096; MUID:95172227; PMID:7867792  
 A:Accession: S69096  
 A:Molecule type: protein  
 A:Residues: 13-37;44-50,'E',52-66 <TR>  
 C:Superfamily: lactose-binding lectin L-36  
 C:Keywords: lectin

Query Match 17.4%; Score 236.5; DB 2; Length 324;  
 Best Local Similarity 35.2%; Pred. No. 1,4e-09;  
 Matches 68; Conservative 25; Mismatches 63; Indels 37; Gaps 10;

QY 71 GADPAGVVPQPSGPGAYBSSGQPSAPGAYATGPYAP-----AGPLI---VPYML 118  
 DB 152 GQPAAGYQGTWIP-AVPSAG-----YMPQNNSLPVVAGPPIENPYPYVG 198  
 QY 119 PLPGGVPRMLITLGTVKPNANRIALDFQGN--DVAHF-PRENNNRVAVCNTKD 175  
 DB 199 TLGGGLARRTIIKGYVLPFAKNLIIINFKVSGTGIAFMNRIQD---CVVRSYNN 254

QY 176 NNWGREERQSV-PFESGKPKICQVLEPDHFKVAVNDAAH-LQYHRYVKLNESKLGIS 233  
 DB 255 GSGWSEERKIPYNPFGAGQFPDLSIRGTRFKVANGQHLPDSHFRFOARFORVDMLEIK 314  
 QY 234 GDIDLTSAASYTKI 246  
 DB 315 GDITL---SVQOI 324

RESULT 12  
 JC6147  
 Prostate carcinoma tumor antigen 1 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 29-Sep-1999  
 C:Accession: JC6147  
 R:Su, Z.Z.; Lin, J.; Shen, R.; Fisher, P.B.; Goldstein, N.I.; Fisher, P.B.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 7252-7257, 1996  
 A:Title: Surface-epitope masking and expression cloning identifies the human prostate c  
 A:Reference number: JC6147; MUID:96293510; PMID:8692978  
 A:Accession: JC6147  
 A:Molecule type: mRNA  
 A:Residues: 1-317 <SUA>  
 A:Cross-references: GB:L78132; NID:g1932711; PIDN:AA851605.1; PID:g1932712  
 C:Comment: This protein is a therapeutic reagent for intervention in perovasive and fata  
 tein is a member of the galectin family.  
 C:Superfamily: lactose-binding lectin L-36  
 C:Keywords: tumor

Query Match 16.5%; Score 223.5; DB 2; Length 317;  
 Best Local Similarity 39.3%; Pred. No. 1,1e-08;  
 Matches 55; Conservative 23; Mismatches 51; Indels 11; Gaps 5;

QY 113 IVPYNLPLPGGVPRMLITLGTVKPNANRIALDFQGN-----DVAHF-PRENNNR 165  
 DB 16 VIFPFGTIPDQDPELTIVIRGHVPSDADRFQVDLQNSSVKPRADVAHFNPFRKAG- 74  
 QY 166 RIVCNTKLDNNWGREERQSVFPESGKPKICQVLEPDHFKVAVNDAAH-LQYHRYVK 224  
 DB 75 -CIVCNTLINERKWEETIYDPFPRKRSFEIIVWLKDKQVAVNGKHTLLYGRIGP- 132  
 QY 225 NEISKLGISGDIIDLTNSASYT 244  
 DB 133 EXIDITLGIYKVNHSIGPS 152

RESULT 13  
 A55975  
 galectin-8 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 29-Sep-1999  
 C:Accession: A55975  
 R:Hadari, Y.R.; Paz, K.; Dekel, R.; Mestrovic, T.; Accili, D.; Zick, Y.  
 J. Biol. Chem. 270, 3447-3453, 1995  
 A:Title: Galectin-8, A new rat lectin, related to galectin-4.  
 A:Reference number: A55975; MUID:9515445; PMID:7852431  
 A:Accession: A55975  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-316 <HAD>  
 A:Cross-references: GB:U09824; NID:g717031; PIDN:AAA66359.1; PID:g717032  
 C:Superfamily: lactose-binding lectin L-36

Query Match 16.1%; Score 218.5; DB 2; Length 316;  
 Best Local Similarity 39.9%; Pred. No. 2,5e-08;  
 Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VPPNPLPGGVPRMLITLGTVKPNANRIALDFQGN-----DVAHF-PRENNNR 166  
 DB 16 IVPVSTITQLRKGLIVIRGHVPSDADRFQVDLQNSSVKPRADVAHFNPFRKSN-- 73  
 QY 167 VIVCNTKLDNNWGREERQSVFPESGKPKICQVLEPDHFKVAVNDAAH-LQYHRYVKLN 225  
 DB 74 CIVCNTLINERKWEETIYDPFPRKRSFEIIVWLKKNFHVAVNGKHTLLYVHRINP-E 132

QY 226 EISKLGISGDIIDLTASV 243  
 DB 133 KIDTLGIFKXNHSIGF 150

RESULT 14  
 155469  
 C:Species: Homo sapiens (man)  
 C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 21-Jul-2000  
 A:Accession: 155469  
 R:Madson, P.; Rasmussen, H.H.; Flint, T.; Gromov, P.; Kruse, T.A.; Honore, B.; Vorum, H.  
 J. Biol. Chem. 270, 5823-5829, 1995  
 A>Title: Cloning, expression, and chromosome mapping of human galectin-7.  
 A:Reference number: 155469; PMID:95197604; PMID:7534301  
 A:Accession: 155469  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-136 <RES>  
 A:Cross-references: GB:I07769; NID:g182131; PIDN:AAA67899.1; PID:g182132  
 C:Genetics:  
 A:Gene: GDB:LCAL57  
 A:Cross-references: GDB:578907; OMIM:600615  
 A:Map position: 19pter-19qter  
 C:Superfamily: beta-galactoside-binding lectin

Query Match 15.5%; Score 211; DB 2; Length 136;  
 Best Local Similarity 35.3%; Pred. No. 3.5e-08;  
 Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;

QY 114 VPYNLPLPGGVVPRMLITITIGVKNANRTALDF---CRGNVAFHF-PRENNRRVY 168  
 DB 4 VPKKSLPGRIGRTGVRLRGVLPNARSFVHLLCGECSALAHFNRDLTSE---V 60

QY 169 VCNTKLDNMGREROSVPPESGKPKIQVLEPDHFKAIVNDA-HLOYNRKKLNEI 227  
 DB 61 VENSKEQSGWGREREPGVDFQRCQFEVLIIASDDGFRAVVDQAYHFRRL-PLAVY 119

QY 228 SKLGISGDIIDLT 240  
 DB 120 RLVEVGVDQLDS 132

RESULT 15  
 CGB07S  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 07-May-1999  
 A:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
 R:Fietzek, P.P.; Allmann, H.; Rautenberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
 A>Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen  
 A:Reference number: A02862; PMID:80026026; PMID:488906  
 A:Accession: A02862  
 A:Molecule type: protein  
 A:Residues: 1-242 <PIB>  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
 A>Title: The covalent structure of calf skin type II collagen. II. The amino acid sequen  
 A:Reference number: A38001; PMID:80026027; PMID:488907  
 A:Accession: A38001  
 A:Molecule type: protein  
 A:Residues: 243-422 <DEMI>  
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
 A>Title: The covalent structure of calf skin type III collagen. III. The amino acid sequen  
 A:Reference number: A38002; PMID:80026028; PMID:488908  
 A:Accession: A38002  
 A:Molecule type: protein  
 A:Residues: 423-571 <BEN>  
 R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979

A>Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequen  
 A:Reference number: A38003; PMID:80026029; PMID:488909  
 A:Accession: A38003  
 A:Molecule type: protein  
 A:Residues: 572-808 <LAN>  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
 A>Title: The covalent structure of calf skin type III collagen. V. The amino acid sequen  
 A:Reference number: A38004; PMID:80026030; PMID:488910  
 A:Accession: A38004  
 A:Molecule type: protein  
 A:Residues: 809-947 <DEMI>  
 R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
 A>Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequen  
 A:Reference number: A38005; PMID:80026031; PMID:488911  
 A:Accession: A38005  
 A:Molecule type: protein  
 A:Residues: 948-1049 <ALT>  
 A:Experimental source: skin  
 R:Henkel, W.  
 Biochem. J. 318, 497-503, 1996  
 A>Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
 A:Reference number: S71946; PMID:96404897; PMID:8809038  
 A:Accession: S71946

A:Molecule type: protein  
 A:Residues: 87-106;1017-1029;1037-1049 <HEN>  
 C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hy  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxyllysine; hydroxyprolin  
 F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>  
 F:1-14/Region: amino-terminal nonhelical telopeptide  
 F:15-1040/Region: helical  
 F:587-589/Region: cell attachment (R-G-D) motif  
 F:752-754/Region: cell attachment (R-G-D) motif  
 F:875-877/Region: cell attachment (R-G-D) motif  
 F:935-937/Region: cell attachment (R-G-D) motif  
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
 F:95,107,119,938,950/Modified site: 5-hydroxylysine (lys) #status experimental  
 F:107,950/Modified site: allysine (lys) #status predicted  
 F:107/Binding site: carbohydrate (lys) (covalent) #status experimental  
 F:1040,1041/Disulfide bonds: interchain #status predicted

Query Match 14.2%; Score 193; DB 1; Length 1049;  
 Best Local Similarity 33.7%; Pred. No. 5.5e-06;  
 Matches 64; Conservative 8; Mismatches 58; Indels 60; Gaps 12;

QY 5 FELHDLG-----SGNPPQGWPGAMQNPAGAGGYGAGYRPG-QAPPGAYPQA 57  
 DB 2 YNAVYKSGVAGGAGGAGVGPAGPPPP--PPGTSGHGA--PGAPGYQGPGR-PQA 56

QY 58 PGAGYHAGPAGY-----PGAPAGVYRPGP--SGPAGY- 89  
 DB 57 GPAGPPGPPGALGPGSGKQSGRGRPRGRPPGPMKMGPRGMGFGPMKGRGRDGN 116

QY 90 -SSGPPSAPGAY---ATGPGAPA--GPLIVPNYMLPLPGVVPMLITITIGTKNNNR 142  
 DB 117 GKSGEPGAPGALGKENGVDGDPGPMGRGAPGRGRPG-----LPGAAGARG- 166

QY 143 IALDFORGND 152  
 DB 167 ---DQARSD 173

RESULT 16  
 S29170  
 C:Species: Mus musculus (house mouse)  
 C>Date: 25-Feb-1994 #sequence\_revision 01-Sep-1995 #text\_change 13-Aug-1999  
 A:Accession: S29170; S46209; S51173

R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.  
 Biochem. J. 289, 735-741, 1993  
 A>Title: Mouse synexin (annexin VII) polymorphisms and a phylogenetic comparison with oc  
 A/Reference number: S29170; MUID:93168121; PMID:7916616  
 A/Accession: S29170  
 A/Molecule type: mRNA  
 A/Residues: 1-463 <ZHA>  
 A/Cross-references: EMBL:U13129  
 R;Zhang-Keck, Z.Y.; Srivastava, M.; Kozak, C.A.; Caohuy, H.; Shirvan, A.; Burns, A.L.; F  
 Biochem. J. 301, 835-845, 1994  
 A>Title: Genomic organization and chromosomal localization of the mouse synexin gene.  
 A/Reference number: S46209; MUID:94330961; PMID:8053909  
 A/Accession: S46209  
 A/Molecule type: DNA  
 A/Residues: 1-463 <ZHF>  
 R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.  
 submitted to the EMBL Data library, May 1993  
 A/Reference number: S51173  
 A/Accession: S51173  
 A/Molecule type: mRNA  
 A/Residues: 1-144, 'S', 146-303, 'A', 305-463 <ZHM>  
 A/Cross-references: EMBL:U13129; NID:9293293; PIND:AAA7238.1; PID:9293294  
 A/Genetics:  
 A:Gene: MGI:Anx7  
 A/Cross-references: MGI:88031  
 A/Map position: 14  
 A/Intons: 19/3; 87/1; 121/1; 142/3; 177/1; 208/3; 246/3; 303/3; 360/3; 386/1; 423/3  
 C/Superfamily: annexin VII; annexin repeat homology  
 C/Keywords: calcium channel; ion channel  
 F;163-234/Domain: annexin repeat homology <AX1>  
 F;235-306/Domain: annexin repeat homology <AX2>  
 F;318-390/Domain: annexin repeat homology <AX3>  
 F;394-463/Domain: annexin repeat homology <AX4>

Query Match 14.2%; Score 192.5; DB 2; Length 463;  
 Best Local Similarity 33.0%; Pred. No. 2.5e-06;  
 Matches 74; Conservative 16; Mismatches 77; Indels 57; Gaps 14;

```

QY 11 LSGSNPNQGMGAMGNPAG-----AGVPGASVYGYGQAP---GAYPGAPG 60
DB 1 MSYPPGP-PTGYPPPPGYPPACQESSFPTAGQTP-----YPSGPPDMGGMAYP-PABSG 52
QY 61 AHGAPGAYP---GAPAPGVYPG--PPSGPAGYPS-----SGQPSAPG-AYA 101
DB 53 GYPPGA-GGYAPAGGYAPAGGYGALSPGPPAPVPGGQGGAPPGAGGSGFYGPQPPQSYG 111
QY 102 TGPY-----GAPAGPLVYPNLPFGVVP-----FMLITLGVKNNARKIALDFQR 149
DB 112 GGPAGVVPVPGFPGGQMPGQY---FGGAPYPSQPAAMTQGTGTLTASN---FDAMR 164
QY 150 GNDVAFHFRFNNNNRNVICNTKLDNNNGRERQSVFPFESGK 193
DB 165 DAEILRRKAMKGFGTDEQALIVDVVNSRNDQRIKAFKTMGK 208

```

RESULT 17  
 S65780  
 glycine/proline-rich protein GPR - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 21-Jul-2000  
 C/Accession: S65780  
 R;Marly, I.; Montfort, A.; Stiefel, V.; Ludevid, D.; Delserny, M.; Puigdomenech, P.  
 Plant Mol. Biol. 30, 625-636, 1996  
 A>Title: Molecular characterization of the gene coding for GPR, a class of proteins ric  
 A/Reference number: S65780; MUID:96189273; PMID:8605310  
 A/Accession: S65780  
 A/Molecule type: DNA  
 A/Residues: 1-177 <MAR>  
 A/Cross-references: EMBL:X84315; NID:9886425; PIND:CA559059.1; PID:91455364  
 A/Note: mRNA was also sequenced  
 C/Genetics:  
 A:Gene: GPRP  
 A/Intons: 99/1

Query Match 14.0%; Score 190.5; DB 2; Length 177;  
 Best Local Similarity 44.6%; Pred. No. 1.2e-06;  
 Matches 50; Conservative 4; Mismatches 41; Indels 17; Gaps 6;

```

QY 13 GSGNPNQGMGAMGNPAG---AGVPGASYP--GYPGAPRAGYQGA--PPGAYHGA 66
DB 2 GGDNDNDKD-KGFHYGPAGYPPGAPPPAGYPPQGGYP--PPGAYPPAGYPPGAYPPAP 58
QY 67 GAYPGAPAPGYPPGPPS-----GPGAYPSGQPSAPGAYATGYPGAPAG 110
DB 59 GGYPPAPRGYGPAPRGYGGYPPAPRGHGYPPAGYPPAHSHGAGIGGMING 110

```

RESULT 18  
 T32783  
 hypothetical protein C50D2.4 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 15-Sep-2003  
 C/Accession: T32783  
 R;Sammons, L.; Woldmann, P.; Bauer, C.  
 submitted to the EMBL Data Library, December 1997  
 A/Description: The sequence of C. elegans cosmid C50D2.  
 A/Reference number: Z21224  
 A/Accession: T32783  
 A/Molecule type: DNA  
 A/Status: preliminary; translated from GB/EMBL/DDBJ  
 A/Residues: 1-329 <SAM>  
 A/Cross-references: EMBL:AF040642; PIND:AMB94952.1; GSPDB:GN00020; CESP:C50D2.4  
 A/Experimental source: strain Bristol N2; clone C50D2  
 A/Genetics:  
 A:Gene: CESP:C50D2.4  
 A/Map position: 2  
 A/Intons: 27/1; 59/3; 76/3; 237/1

Query Match 14.0%; Score 189.5; DB 2; Length 329;  
 Best Local Similarity 43.4%; Pred. No. 2.8e-06;  
 Matches 53; Conservative 1; Mismatches 45; Indels 23; Gaps 7;

```

QY 19 PGWPGAMGNQ-PAGAGGYPGASVYGP-----QAPPGAYPQAPPGA-----YHGAPGA 68
DB 187 PTGPPGPPKQKQPKPNRPERGA--FGRGGMALPGRPPGPPGQGGADGLPGNPGAPGV 244
QY 69 -----YGPAPAPGYPPGPPSGPAGYPSGQPSAPGAYATGYGAPAGLI--VPYNLPL 120
DB 245 PGQVIEVPTGPAPA---GPPGPPGPIGPPGQGGAGGSSQPGGQGGDDGIDGAPGNPS 301
QY 121 PG 122
DB 302 PG 303

```

RESULT 19  
 T37216  
 beta-galactoside-binding protein GBP - Caenorhabditis elegans  
 N/Alternate names: beta-galactoside-binding lectin homolog, 32K  
 C/Species: Caenorhabditis elegans  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
 C/Accession: T37216; T26324; A42846; S27798  
 R;Arata, Y.; Hirabayashi, J.; Kasai, K.  
 J. Biol. Chem. 272, 26669-26677, 1997  
 A>Title: Structure of the 32-kDa galectin gene of the nematode Caenorhabditis elegans.  
 A/Reference number: Z21635; MUID:97476274; PMID:9334250  
 A/Accession: T37216  
 A/Status: preliminary; translated from GB/EMBL/DDBJ  
 A/Molecule type: DNA  
 A/Residues: 1-279 <ARA>  
 A/Cross-references: EMBL:AB008082; NID:92564035; PIND:BA22942.1; PID:92564036  
 R;Smey, R.  
 submitted to the EMBL Data Library, November 1996  
 A/Reference number: Z20198  
 A/Accession: T26324  
 A/Status: preliminary; translated from GB/EMBL/DDBJ

	Query Match	13.9%	Score 188.5;	DB 2;	Length 279;
	Best Local Similarity	36.4%	Pred. No. 2.8e-06;		
	Matches	47;	Conservative	26;	Mismatches 49; Indels 7; Gaps 5.
QY	114 VPYNLPDGVVPRMLITLGTVEKNARIALDFQRGN-DVAFHF--PRENNENRRIAYCN	171			
Dd	150 VPPESGLANGLPVGKSLVFGEVEKKARKFHNLIRKGKDLSFFHNPFDDEGH---VRN	206			
QY	172 TKLDNNNGREERÖSVFPPESSGPFKEIQVLVEEDHKFAVN--DAHLQYHHRYVKLNLEISKI	230			
Dd	207 SLANNEMWENBEREGNIPFEKVGQGLVTIQNEEYANQÖVFNBERIRYSISFAHRADP-HDIAGL	265			
QY	231 GISGDIDLT 239				
Dd	266 QISGDIELS 274				

Query Match	13.9%	Score 188.5;	DB 2;	Length 285;
Best Local Similarity	36.4%	Pred. No. 2.8e-06;		
Matches 47;	Conservative 26;	Mismatches 49;	Indels 7;	Gaps 5

Query Match	13.9%	Score 188	DB 2	Length 744	
Best Local Similarity	30.5%	Pred. No. 8.6e-06			
Matches	60	Conservative	12	Mismatches	67
				Indels	58
				Gaps	9
QY	15	GNPNPGCMPGAMG---	NQAPAGAGYPG---	ASYPGYPG-----	46
Db	429	GEPLGQGFPPKRGFLGEVGPCKMKGFP	PLGPKGHEGKGKVGPLPGVFGLLGPKKEP	GLP	488
QY	47	-----QAPPGAYPQAPPPGAYHAGP---	AYPGAPPGYF-----	GPSPGP	85
Db	489	GDQGLQGPFG--IPGIIGPSGPIGPBGL	PGPGEGLGPPPGPFGIGKGVAGLHGP	PEPKP	547
QY	86	GAYESSGQSPAPGAYATGPGAPGAPLIV	PNLPLPGVAVPRMLITLIGATKPNAN	IAL	145
Db	548	GALGPQGGPGLPG--PPGP	PGPPPPAVMPTTTPPGQBYLPLDKL	GLDGVAFPHAT-----	601
QY	146	DFQRGND--VAHFPPPT	160		
Db	602	GAKKXKNGGPAWEMPAF	618		

Query Match	13.84	Score 187.5	DB 2	Length 313
Best Local Similarity	43.98	Pred. No. 3,7e-06		
Matches 50	Conservative 2	Mismatches 41	Indels 21	Gaps 6

QY 14 SGNPFGWGMGNQAGAGYF---GASYGPGQAPPGAYPGQAPGAYHAGAP- 68  
 DB 172 AGPPGPGAPQKQ--PSGAPGVGQGTGASLPPEPPAPGPPSGGPGSGNAGAPGAP 229  
 QY 69 ----YCGAPAPGVYPPGPPSGPAGYF-----SSGQPSAPGAVATGPGAPAGP 111  
 DB 230 GQYVDVGTGTGPPAPGPPPP- GPACAPQPGSSSGGQGGGCG--PQGNAGAPGAP 280

## RESULT 23

A85217  
 hypothetical protein AT4g19200 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: A85217  
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; MUID:20083488; PMID:10617138  
 A:Accession: A85217  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-179 <STO>  
 A:Cross-references: GB:NC\_001268; NID:97268715; PIDN:CAB78922.1; GSPDB:GN00140  
 C:Genetics:  
 A:Gene: AT4g19200  
 A:Map position: 4

Query Match 13.7%; Score 186.5; DB 2; Length 179;  
 Best Local Similarity 45.7%; Pred. No. 2.4e-06;  
 Matches 43; Conservative 5; Mismatches 31; Indels 15; Gaps 5;

QY 21 GMPGAMGNQAPGAGYPGASYPGYPGQAPPGAYPGQA-PPGAYHAGAPGAYHAGAPGYP 79  
 DB 18 GPPGG-CHYPPAGQGGYPPQGGYPPQGGYPPAGYPPAGYPPGAYPPGAYPPAGYPP-PAFGYYP 74  
 QY 80 GPPSGGAYPPSSGQPSAPGAVATGPGYAPAGPLI 113  
 DB 75 ----PAGYF-----APGAMHSHSGGGLGAMI 97

## RESULT 24

T04441  
 hypothetical protein T18B16.170 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
 C:Accession: T04441  
 R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoerge, W.; Bancroft, I.; Mewes, H.W.  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: Z15359  
 A:Accession: T04441  
 A:Molecule type: DNA  
 A:Residues: 1-277 <BEV>  
 A:Cross-references: EMBL:AL021687  
 A:Experimental source: cultivar Columbia; BAC clone T18B16  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 87/1; 144/1; 179/1; 207/3  
 A>Note: T18B16.170

Query Match 13.7%; Score 186.5; DB 2; Length 277;  
 Best Local Similarity 45.7%; Pred. No. 3.8e-06;  
 Matches 43; Conservative 5; Mismatches 31; Indels 15; Gaps 5;

QY 21 GMPGAMGNQAPGAGYPGASYPGYPGQAPPGAYPGQA-PPGAYHAGAPGAYHAGAPGYP 79  
 DB 18 GPPGG-CHYPPAGQGGYPPQGGYPPQGGYPPAGYPPAGYPPGAYPPGAYPPAGYPP-PAFGYYP 74  
 QY 80 GPPSGGAYPPSSGQPSAPGAVATGPGYAPAGPLI 113  
 DB 75 ----PAGYF-----APGAMHSHSGGGLGAMI 97

## RESULT 25

LNUH7  
 annexin VII, long form - human  
 N:Alternate names: synexin  
 N:Contents: annexin VII, long form; annexin VII, short form  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Sep-1992 #sequence\_revision 26-Jan-1996 #text\_change 22-Jun-1999  
 C:Accession: A54467; A32554; A39513; B39513  
 R:Shirvan, A.; Srivastava, M.; Wang, M.G.; Cultraro, C.; Magendzo, K.; McBride, O.W.; P.  
 Biochemistry 33, 6888-6901, 1994  
 A:Title: Divergent structure of the human synexin (annexin VII) gene and assignment to C  
 A:Reference number: A54467; MUID:94264005; PMID:7515686  
 A:Accession: A54467  
 A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA  
 A:Residues: 1-488 <SHI>  
 R:Burns, A.L.; Magendzo, K.; Shirvan, A.; Srivastava, M.; Rojas, E.; Alijani, M.R.; Pol  
 Proc. Natl. Acad. Sci. U.S.A. 86, 3798-3802, 1989  
 A:Title: Calcium channel activity of purified human synexin and structure of the human  
 A:Reference number: A32554; MUID:89264510; PMID:2542947  
 A:Accession: A32554  
 A:Molecule type: mRNA  
 A:Residues: 1-145,168-488 <BUR>  
 A:Cross-references: EMBL:J05453; NID:9338243; PIDN:AAA36616.1; PTD:9338244  
 R:Magendzo, K.; Shirvan, A.; Cultraro, C.; Srivastava, M.; Pollard, H.B.; Burns, A.L.  
 J. Biol. Chem. 266, 3228-3232, 1991  
 A:Title: Alternative splicing of human synexin mRNA in brain, cardiac, and skeletal mus  
 A:Reference number: A39513; MUID:91131630; PMID:1825209  
 A:Accession: A39513  
 A:Molecule type: mRNA  
 A:Residues: 137-145,168-176 <MAG>  
 A:Cross-references: EMBL:J05732  
 A:Accession: B39513

A:Molecule type: mRNA  
 A:Residues: 137-176 <MA2>  
 C:Comment: Annexin undergo reversible, calcium-dependent binding to membrane phospholi  
 derolipid.

C:Comment: The long form of annexin VII is more prevalent in brain, heart, and skeletal  
 C:Genetics:  
 A:Gene: GDB:ANX7  
 A:Cross-references: GDB:369042; OMIM:186360  
 A:Map position: 10q21.1-10q21.2  
 A:Introns: 17/3; 87/1; 124/1; 145/3; 167/3; 202/2; 233/3; 271/3; 328/3; 385/3; 411/1; 4  
 C:Superfamily: annexin VII; annexin repeat homology  
 C:Keywords: alternative splicing; calcium binding; calcium channel; duplication; endome  
 F:1-488/Product: annexin VII, long form #status predicted <MA1>  
 F:1-145,168-488/Product: annexin VII, short form #status predicted <MA3>  
 F:1188-259/Domain: annexin repeat homology <AX1>  
 F:1199-215/Region: endoneixin fold #status predicted  
 F:260-331/Domain: annexin repeat homology <AX2>  
 F:271-287/Region: endoneixin fold #status predicted  
 F:343-415/Domain: annexin repeat homology <AX3>  
 F:355-371/Region: endoneixin fold #status predicted  
 F:419-488/Domain: annexin repeat homology <AX4>  
 F:430-446/Region: endoneixin fold #status predicted

Query Match 13.7%; Score 186; DB 1; Length 488;  
 Best Local Similarity 41.4%; Pred. No. 7.5e-06;  
 Matches 48; Conservative 2; Mismatches 44; Indels 22; Gaps 6;

QY 17 PNPQGMGAMGNQAPGAGYPGASYPGYPGQAPPGAYPGQA-PPGAYHAGAPGAYHAGAPG 76  
 DB 33 PVPGGP-----PMGCGAYPQVPSSGYDG---AGGYPA---PGGY-PAFGYPPAGPQ 79  
 QY 77 VVPGGPSSGAYPPSSGQPSAPGAVATGPGYAPAGPLI 126  
 DB 80 ----GASYPGVPPGGGFGVPPGAGGPGYPPQPPSGSYGGGPAQVPLPAGGPGGMP 132

## RESULT 26

A27353  
 collagen alpha 1(III) chain precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)  
C>Date: 28-Aug-1999 #sequence\_revision 30-Sep-1990 #text\_change 13-Aug-1999  
C/Accession: A27353; A22287; P50066  
R/Mood, L.; Theriault, N.; Vogeli, G.  
Gene 61, 225-230, 1987  
A>Title: Complete nucleotide sequence of the N-terminal domains of the murine alpha-1  
A/Reference number: A91588; MUID:86167858; PMID:3443309  
A/Accession: A27353  
A/Molecule type: mRNA  
A/Residues: 1-488 <MOO>  
A/Cross-references: GB:M18933; NID:9192276; PIDN:AAA37338.1; PID:g553983  
A/Note: in the helical region, Gly-X-Y repeat is interrupted by 287-Val instead of Gly  
R/Lian, G.; Mody, M., de Crombrugne, B.  
J. Biol. Chem. 260, 3773-3777, 1985  
A>Title: Identification of the promoter and first exon of the mouse alpha 1(III) collag  
A/Reference number: A92513; MUID:85131189; PMID:33972847  
A/Accession: A22287  
A/Molecule type: DNA  
A/Residues: 1-28 <LTA>  
C/Genetics: 29/1  
A/Intons: 29/1  
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C/Keywords: coiled coil; extracellular matrix; glycoprotein; sulfoprotein; trimer; trip  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-154/Domain: amino-terminal propeptide #status predicted <APR>  
F:155-488/Product: von Willebrand factor type C repeat homology <WVC>  
F:164-488/Region: helical (fragment)  
F:40,152,159/Binding site: sulfate (Tyr) (covalent) #status predicted

Query Match 13.7%; Score 185.5; DB 2; Length 488;  
Best Local Similarity 36.5%; Pred. No. 8.2e-06;  
Matches 57; Conservative 10; Mismatches 46; Indels 43; Gaps 10;

QY 5 FSLHDLISG-----SGNPPOGMPGAMGNOPAGAGVPSGSPYPS-QAPPGAYPGQAP 59  
DB 156 FSYDYKSGVGMAGVPGAPGPPGLPG--PPSSGHPGS--PSSPYQGPFGP-PQQAQP 210  
QY 60 GAYHGAPGAY-----PGAPAPGVYPPGP--SGPAGVPSGSPAPAGAYA-- 101  
DB 211 AGPPGPPGALPGAPGAGKDGESGRPGRPGRGLPPGPIKGP-----AGMPGPPGKGR 265  
QY 102 -----TGPYGAAP--GELIVPYNLPLPGGVNPR 127  
DB 266 GPDGRNGEKETGAPGLKGENVLPQDNGAPGPMGPR 301

## RESULT 27

T19564  
Hypothetical protein C29F4.1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T19564  
R/Kershaw, J.  
submitted to the EMBL Data Library, January 1996  
A/Reference number: Z19143  
A/Accession: T19564  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-299 <WIL>  
A/Cross-references: EMBL:Z68335; PIDN:CAA92729.1; GSPDB:GN00022; CESP:C29F4.1  
A/Experimental source: clone C29F4  
C/Genetics:  
A/Gene: CESP:C29F4.1  
A/Map position: 4  
A/Intons: 27/3

Query Match 13.6%; Score 184; DB 2; Length 299;  
Best Local Similarity 42.9%; Pred. No. 6.2e-06;  
Matches 54; Conservative 4; Mismatches 46; Indels 22; Gaps 8;

DB 172 GARGQDQPPAPKNGSGPNKPNBAPGAPGQPGQDAPSPPIPPGA-PGPAPGPGQGP 230  
QY 67 GA-----YGPAPGVYPPGPPSGPSSGSPAPGAYATGYPGAPPLIVPNL 118  
DB 231 GARGQPHDQGPAPGP-----KPNGNPNBAPGADGNPAPGQAGTAGAGKNG--ICPKYC 265  
QY 119 PLPGV 124  
DB 286 ALDGGV 291

## RESULT 28

S23779  
collagen alpha 1(VIII) chain - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C/Accession: S23779  
R/Muragaki, Y.; Shiota, C.; Inoue, M.; Ooshima, A.; Olsen, B.R.; Ninomiya, Y.  
Eur. J. Biochem. 207, 895-902, 1992  
A>Title: Alpha-1(VIII)-collagen gene transcripts encode a short-chain collagen polypepti  
A/Reference number: S23779; MUID:92362626; PMID:1499564  
A/Accession: S23779  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-743 <WUR>  
A/Cross-references: EMBL:X66976; NID:g50493; PIDN:CAA47387.1; PID:g1359953  
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
F:16-742/Domain: complement C1q carboxyl-terminal homology <CIQ>

Query Match 13.5%; Score 183.5; DB 1; Length 743;  
Best Local Similarity 34.6%; Pred. No. 1.8e-05;  
Matches 55; Conservative 8; Mismatches 57; Indels 39; Gaps 8;

QY 15 GNPPOGMPGAMGNOPAGAGVYG-ASYGVYPS-----QAPPGAYPGQA 57  
DB 447 GPPGMRGALPPIG--PKGGHKGHLPLGVYGLGPKKEPGLPDQGLQGPFG-IPGLV 503  
QY 58 PPGAYHGAPG-----AYPAPAPGVY-----GPPSGPAGVPSGSPAPAGAYA 101  
DB 504 GPGSPGPPGIPGYPGEPGLPPGPPGPGVGRKPVAGLHGPBPKPAGLPQGPGLPG--P 561  
QY 102 TGPYGAAPGELIVPYNLPLPGGVNPRMLITLIGTVKPN 140  
DB 562 PGPDPGPPAPVMPPIPSF--QGEYLPDMGLIDGVTPPIA 599

## RESULT 29

T25407  
Hypothetical protein T28C6.4 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T25407  
R/Lloyd, C.  
submitted to the EMBL Data Library, September 1995  
A/Reference number: Z20030  
A/Accession: T25407  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-299 <WIL>  
A/Cross-references: EMBL:Z54238; PIDN:CAA90995.1; GSPDB:GN00022; CESP:T28C6.4  
A/Experimental source: clone T28C6  
A/Accession: T25409  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-299 <WIL>  
A/Cross-references: EMBL:Z54238; PIDN:CAA90997.1; GSPDB:GN00022; CESP:T28C6.6  
C/Genetics:  
A/Gene: CESP:T28C6.4; CESP:T28C6.6  
A/Map position: 4  
A/Intons: 27/3

Query Match 13.5%; Score 183; DB 2; Length 299;



Best Local Similarity 42.9%; Pred. No. 7.3e-06;  
Matches 54; Conservative 4; Mismatches 46; Indels 22; Gaps 8;  
QY 15 GNPNGQMGWANGN-PAGAGGYPGA-SYGYPGQ-AP-----PGAYPGQAPPGAYHGAP 66  
Db 172 GAPGNDGQPGAKNKPSPSPNPGAPGAPGQPGDAPSEPIITPGA-PGAPGAPGQGP 230  
QY 67 GA-----YPGAPAPGYPPGPPSGPAGYPPSSGQSPAGAYATGPGYAPAGPLIVPNL 118  
Db 231 GAPGCGHGDGQGPAPG---KGPNGNPGAPGADGNPGAPGQAGTSGAGEKG--ICPKYC 285  
QY 119 PLPGV 124  
Db 286 AIDGV 291

RESULT 30  
T23960  
hypothetical protein F36A4.6 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T23960  
R/Miller, N.; Bradshaw, H.  
submitted to the EMBL Data Library, April 1996  
A/Description: The sequence of C. elegans cosmid F36A4.  
A/Reference number: Z20713  
A/Accession: T23960  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-317 <MIL>  
A/Cross-references: EMBL:U53333; PIDN:MAA96159.1; GSPDB:GN00022; CESP:F36A4.6  
A/Experimental source: strain Bristol N2; clone F36A4  
C/Genetics:  
A/Gene: CESP:F36A4.6  
A/Map position: 4  
A/Introns: 9/3; 43/3; 70/3

Query Match 13.4%; Score 182.5; DB 2; Length 317;  
Best Local Similarity 43.6%; Pred. No. 8.4e-06;  
Matches 51; Conservative 3; Mismatches 44; Indels 19; Gaps 7;  
QY 15 GNPNGQMGWANGN-PAGAGGYPGA-SYGYPGQ-APPGAYPPGQAPPGAYHGAP 67  
Db 205 GPPGPPGPGA-----PBPGRPGDAPSEPIITPBPBPQGPGR-PGQAGPDGQPGAPG 258  
QY 68 AYPGAPAPGYPPGPPSGPAGYPPSSGQSPAGAYATGPGYAPAGPLIVPNLPLPGV 124  
Db 259 G-PGTPGS---KGPNGNPGAPGADGNPGAPG--QAGPAGAKGKGIKPKYCAIDGV 309

RESULT 31  
T26185  
hypothetical protein W05B2.1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T26185  
R/Gardner, A.  
submitted to the EMBL Data Library, October 1996  
A/Reference number: Z20166  
A/Accession: T26185  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-304 <MIL>  
A/Cross-references: EMBL:Z81138; PIDN:CAB03475.1; GSPDB:GN00021; CESP:W05B2.1  
A/Experimental source: clone W05B2  
C/Genetics:  
A/Gene: CESP:W05B2.1  
A/Map position: 3  
A/Introns: 27/3

Query Match 13.4%; Score 181.5; DB 2; Length 304;  
Best Local Similarity 44.8%; Pred. No. 9.4e-06;  
Matches 56; Conservative 6; Mismatches 42; Indels 21; Gaps 9;

QY 12 SGSGNP---NPQMGWANGN-PAGAGGYPGA-SYGYPGQAPPGAYPGQAPPGAY 62  
Db 180 SGAGAPGAPGKAGAPGAGNPGQAPGQPGSDAQSSPPGAPGQGP-----QGPGR- 233  
QY 63 HGAPATGAPAPGYPPG--PSG-PGAYPPSSGQSPAGAYATGPGYAPAGPLIVPNL 119  
Db 224 AGSPGA-PGGPGQAGAPGPKPSGAPGQPGADGNPGAGC--QPGSGGAGKGIKPKYCA 290  
QY 120 LPGV 124  
Db 291 IDGV 295

RESULT 32  
T26184  
hypothetical protein W05B2.6 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T26184  
R/Gardner, A.  
submitted to the EMBL Data Library, October 1996  
A/Reference number: Z20166  
A/Accession: T26184  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-304 <MIL>  
A/Cross-references: EMBL:Z81138; PIDN:CAB03474.1; GSPDB:GN00021; CESP:W05B2.6  
A/Experimental source: clone W05B2  
C/Genetics:  
A/Gene: CESP:W05B2.6  
A/Map position: 3  
A/Introns: 27/3

Query Match 13.4%; Score 181.5; DB 2; Length 304;  
Best Local Similarity 44.8%; Pred. No. 9.4e-06;  
Matches 56; Conservative 6; Mismatches 42; Indels 21; Gaps 9;  
QY 12 SGSGNP---NPQMGWANGN-PAGAGGYPGA-SYGYPGQAPPGAYPGQAPPGAY 62  
Db 180 SGAGAPGAPGKAGAPGAGNPGQAPGQPGSDAQSSPPGAPGQGP-----QGPGR- 233  
QY 63 HGAPATGAPAPGYPPG--PSG-PGAYPPSSGQSPAGAYATGPGYAPAGPLIVPNL 119  
Db 224 AGSPGA-PGGPGQAGAPGPKPSGAPGQPGADGNPGAGC--QPGSGGAGKGIKPKYCA 290  
QY 120 LPGV 124  
Db 291 IDGV 295

RESULT 33  
T22827  
hypothetical protein F57B1.4 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T22827  
R/Sims, M.  
submitted to the EMBL Data Library, August 1996  
A/Reference number: Z19622  
A/Accession: T22827  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-356 <MIL>  
A/Cross-references: EMBL:Z78064; PIDN:CAB01508.1; GSPDB:GN00023; CESP:F57B1.4  
A/Experimental source: clone F57B1  
C/Genetics:  
A/Gene: CESP:F57B1.4  
A/Map position: 5  
A/Introns: 65/1; 111/3

Query Match 13.3%; Score 180.5; DB 2; Length 356;  
Best Local Similarity 43.0%; Pred. No. 1.3e-05;

collagen col-13 precursor - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans







QY 13 GSGN---PNQGMFGAMNQ-PAGAGGYFGASYPGYFGAPGAYPGQAPFG--AYHGAP 66  
 Db 134 GNGNPGSPGGDGGPFGAPGKMGKPSGPMGNFGA--FGAPG-----PGGDADSEPTTGPAP 185  
 QY 67 GAYFGAPAPGVYGRPPSGFGAYRSSQSPAPGAYVTGFGYAPAGFLI-----VEYNLFLP 121  
 Db 186 GR-QGTTGR---QGPRGGPGGPGHDDGQAPFG--FKGNGNPGGGGAGNAGNFGAGGSGTTP 239  
 QY 122 GGV 124  
 Db 240 GGV 242

RESULT 49

1.22/05  
hypothetical protein F55C10.2 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 15-Sep-2003  
C:Accession: T22705  
R:Dobson, R.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z19603  
A:Accession: T22705  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-239 <MIL>  
A:Cross-references: EMBL:Z74036; PIRN:CAA9486.1; GSPDB:GN00023; CBSP:F55C10.2  
C:Experimental source: clone F55C10  
C:Genetics:  
A:Gene: CBSP:F55C10.2  
A:Map position: 5  
A:Introns: 54/3

Query Match	12.8%	Score 173.5	DB 2	Length 299
Best Local Similarity	43.1%	Pred. No. 3.3e-05		
Matches 53	Conservative 3	Mismatches 42	Indels 25	Gaps 9

[illegible]

## RESULT 50

S70644  
 annexin VII - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 13-Aug-1999  
 C:Accession: S70644  
 R:Stratovaara, M.; Zhang-Keck, Z.Y.; Caohuy, H.; McPhie, P.; Pollard, H.B.  
 Biochem. J. 316, 729-735, 1996  
 A>Title: Novel isoforms of synexin in Xenopus laevis: multiple tandem PGQM repeats distributed  
 A:Reference number: S70644; MUID:96265035; PMID:8670145  
 A:Accession: S70644  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-512 <SRI>  
 A:Cross-References: EMBL:U016365; NID:g790543; PIDD:AAB1845.1; PID:g790544  
 C:Superfamily: annexin VII; annexin repeat homology  
 F:212-282/Domain: annexin repeat homology <X1>  
 F:283-355/Domain: annexin repeat homology <X2>  
 F:367-439/Domain: annexin repeat homology <X3>  
 F:443-512/Domain: annexin repeat homology <X4>

```

Best Local Similarity 31.6%; Pred. No. 5.9e-05;
Matches 73; Conservative 16; Mismatches 71; Indels 71; Gaps 17;

QY 12 SSGSNPNQ--GWPAGMGNCPAGAGGYGA--SYGYG-----QAPPGA---Y 53
Db 48 SGGGNAPPPGGGYPIA-GGYA-PGYDGMGMRTPGAPGAPAGCGGYGAPGAPAYGV 105
QY 54 PGQAPGAYIHAPGAYPGAPAPGYVPGPPSG---PGAYPSSG--QPS----- 95
Db 106 PGYGGPG-FNAPAGGY-GAPNAGGFGVPPAGGYSGPGAPGYGGFSPSPSSQSYAGGPGQ 163
QY 96 -----APGAYATGPGYAGPAGPLVYVLLPLPGGYVRMLITLTGVENNRRI 143
Db 164 MPGGMPGMPGQAPSGTSPSG-APAPD--TPY-----AAMATGQTIKAEVFD 210
QY 144 ALDPQRGNDAVHFHPFNENNRV-IVCNKTLDNNGREERQSYVPPESGK 193
Db 211 AL--SDAEKLRKAMKGFIDEKPLDVVANSNQ--RQIKQAAPKAYGK 256

Search completed: August 23, 2004, 14:38:57
Job time : 25 secs

```

Search completed: August 23, 2004, 14:38:57  
Job time : 25 secs

Query Match 12.8%; Score 173.5; DB 2; Length 512;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:32:27 ; Search time 52 Seconds  
(without alignments)  
1492.644 Million cell updates/sec

Title: US-09-297-040-4  
Perfect score: 1357  
Sequence: 1 MAUNFSLHDLALSGSGNPNPQ.....ISKLGISGIDILTSASYTMI 246

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 65 summaries

Database :  
1: SP\_TREMBL\_25:\*  
2: SP\_Archaea:\*  
3: SP\_Bacteria:\*  
4: SP\_Fungi:\*  
5: SP\_Invertebrate:\*  
6: SP\_Mammal:\*  
7: SP\_Misc:\*  
8: SP\_Organelle:\*  
9: SP\_Phage:\*  
10: SP\_Plant:\*  
11: SP\_Rodent:\*  
12: SP\_Virus:\*  
13: SP\_Vertebrate:\*  
14: SP\_Unclassified:\*  
15: SP\_Virus:\*  
16: SP\_Bacteriap:\*  
17: SP\_Archivep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1082	79.7	264	11	Q8C253
2	695.5	51.3	326	13	Q90713
3	663.5	48.9	332	13	Q8QGD9
4	603	44.4	121	4	Q86T75
5	581	42.8	108	4	Q81XB9
6	580.5	42.8	261	13	Q7ZSY1
7	444	32.7	139	11	Q61357
8	314.5	23.2	341	13	Q9P111
9	299	22.0	322	11	Q99183
10	284	20.9	320	13	Q7T119
11	271	20.0	348	13	Q8QGB1
12	269.5	19.9	310	13	Q7T118
13	264	19.5	311	4	Q8WY07
14	260	19.2	349	6	Q9XSM9
15	258.5	19.0	315	13	Q7ZT85
16	258	19.0	343	13	Q8W998

17	257	18.9	317	6	Q9XSM8	Q9xsm8 sus scrofa
18	246.5	18.2	340	13	Q8W999	Q8w999 xenopus lae
19	243.5	17.9	326	11	Q91X74	Q91x74 mus musculus
20	238.5	17.6	328	13	Q7ZTB9	Q7ztb9 xenopus lae
21	237.5	17.5	326	11	Q8K419	Q8k419 mus musculus
22	233.5	17.2	328	6	Q9TUB8	Q9tub8 oryctolagus
23	230	16.9	308	13	Q7ZTB8	Q7ztb8 xenopus lae
24	228.5	16.8	359	4	Q8TEV1	Q8tev1 homo sapien
25	223.5	16.5	359	4	Q8BXC8	Q8bxc8 homo sapien
26	219.5	16.2	204	13	Q7ZYP0	Q7zyp0 xenopus lae
27	219.5	16.2	319	13	Q7ZTB6	Q7ztb6 xenopus lae
28	216.5	16.0	300	11	Q88353	Q88353 mus musculus
29	215.5	15.9	180	13	Q7T120	Q7t120 brachydanto
30	213	15.7	332	13	Q8UW97	Q8uw97 xenopus lae
31	210.5	15.5	703	12	Q83467	Q83467 porcine ade
32	206.5	15.2	162	6	Q8MNS9	Q8mns9 ovis aries
33	195.5	14.4	903	16	Q8ZHF3	Q8zhf3 streptomyce
34	195	14.4	605	13	Q801J8	Q801j8 salmo salar
35	194	14.3	223	11	Q8C6H0	Q8c6h0 mus musculus
36	192.5	14.2	415	11	Q8CCV9	Q8ccv9 mus musculus
37	192.5	14.2	463	11	Q922A2	Q922a2 mus musculus
38	192.5	14.2	463	11	Q8BP75	Q8bp75 mus musculus
39	192.5	14.2	485	13	Q804G3	Q804g3 brachydanto
40	191	14.1	483	13	Q7391	Q7391 brachydanto
41	190.5	14.0	177	10	Q39115	Q39115 aradidopsis
42	189.5	14.0	329	5	Q44796	Q44796 caenorhabdi
43	189.5	14.0	526	13	Q804G4	Q804g4 brachydanto
44	188.5	13.9	285	5	Q45904	Q45904 caenorhabdi
45	187.5	13.8	277	5	Q9NV1	Q9nv1 haemochus
46	187.5	13.8	283	5	Q9NV0	Q9nv0 haemochus
47	187.5	13.8	313	5	Q20922	Q20922 caenorhabdi
48	187	13.8	300	4	Q96Q87	Q96q87 homo sapien
49	186.5	13.7	179	10	Q9M0L8	Q9m0l8 aradidopsis
50	186.5	13.7	277	10	Q43678	Q43678 aradidopsis
51	185.5	13.7	463	11	Q8VIN2	Q8vin2 ratcus norv
52	185	13.6	185	10	Q84TC1	Q84tc1 oryza sativ
53	185	13.6	1070	2	Q9APM8	Q9apm8 myxococcus
54	184	13.6	239	5	Q18302	Q18302 caenorhabdi
55	183.5	13.5	744	11	Q921S8	Q921s8 mus musculus
56	183.5	13.5	744	11	Q8BGL6	Q8bgl6 mus musculus
57	183	13.5	239	5	Q27318	Q27318 caenorhabdi
58	183	13.5	1884	5	Q9NFW2	Q9nfw2 nephila ina
59	182.5	13.4	304	5	Q20091	Q20091 caenorhabdi
60	181.5	13.4	304	5	Q9XVG3	Q9xvg3 caenorhabdi
61	181.5	13.4	304	5	Q9U349	Q9u349 caenorhabdi
62	181.5	13.4	304	5	Q9U348	Q9u348 caenorhabdi
63	181.5	13.4	308	5	Q94620	Q94620 meloidogyne
64	181.5	13.4	462	5	Q9NFW3	Q9nfw3 nephila cla
65	181.5	13.4	907	5	Q44359	Q44359 nephila cla

## ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	264 AA.
Q8C253			
ID	Q8C253		
AC	08C253		
DT	01-MAR-2003 (TREMBLrel. 23, Created)		
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Lectin.		
GN	LGALB3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
FM	SEQUENCE FROM N.A.		
RP	STRAIN=NOD;		
RC	MEDLINE=22354683; Pubmed=12466851;		
RX	THE PANTOM Consortium.		
RA	THE RIKEN Genome Exploration Research Group Phase I & II Team;		



```

RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK089249; BAC00813.1; -.
DR PIR; A45983; A45983.
DR MGD; MGI:96778; Lgals3.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005515; F:protein binding; IPT.
DR InterPro; IPR008985; Cona_like_lectin.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal_bind_lectin.1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALACTIN; 1.
DR SEQUENCE 264 AA; 27415 MW; 253AA4F80371A3740 CRC64;

```

Query Match	79.7%;	Score 1082;	DB 11;	Length 264;
Best Local Similarity	78.6%;	Pred. No. 1.2e-65;		
Matches 209;	Conservative 15;	Mismatches 20;	Indels 22;	Gaps 8

2Y 1 MAQNPSLHALSGSGNPNGQMPGAWNGNDAGAGGYPGASYPG-YPGQAPPGAYGQAPP 59

Db 1 MAQSFLNALAGSGNPNGQYPGAWNGND-GAGGYPGAYPGAYPGQAPPGAYPGQAPP 59

QY 60 GAHGG--AGCAYPCGAPPCGYPPGPPSGPCAYPPSS-----GQESAPGAY-----AT 102  
Db 60 GAAPGQAPPSAYPCGTAPGAPYPG--PTAPGAYPPGSGTAPGAPGAPPGQPCAPGAYGAPSAAGYPPAA 118	

QY 103 GPYCAPAGCLITVPYNLPDPGCVVPRMLITLTGLTYKPNANNIALDFQRGNDAFHF--PREN 161

Db 119 GPVCPAGCLITVPYDLPLFGGVMPRLITMTGTVENANRIVLDRRGNVAFHFNPREN 178

QY 162 ENNRVIVCNTKLDNNMGEEERQSVPPFESGKPEFKIQLVLEPDHKKVAANDAH-LQYNHR 220

DB 179 ENNRVIVCNTKDDNNMGKEERQSGPPFESGKPEFKIQVLEADHKKVAANDAHILQYNHR 238

Qy 221 VKKNEISKIGSDIDLTSASYMTI 246  
 : | | | | | | | : | |  
 Db 239 MKNI.PRTSOLIGSDITLTSANHAMT 264  
 : | | | | | | | : | |

RESULT 2  
00073

ID	C90713	PRELIMINARY;	PRT;	262 AA.
AC	C90713:			
DT	01-NOV-1996	(TREMBlrel. 01, Created)		

DT 01-OCT-2003 (TREMBlérel. 25, last annotation update)  
DE Galectin-3.  
OS Gallus gallus (Chicken).

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OC NCBI\_TaxID=9031;  
OX

RA Nurminkaya M.V.; Linsenmayer T.F.;  
RC TISSUE=Cartilage;  
RP SEQUENCE FROM N.A.  
BP

RT chondrocyte hypertrophy." ;  
 RL Dev. Dyn. 0:0-0 (1996).  
 DR EMBL: U50339; AAB02856.1 ; -

DR HSBV; PI/93.1; IAKK.  
DR InterPro; IPR008985; ConA\_like\_1ec\_g1.  
DR InterPro; IPR001079; Galectin.  
DR Pfam; PF00337; Gal-binding lectin; 1

```
DR SMART; SM00276; GJECT; 1.  
DR PROSITE; PSC0309; GALACTIN; 1.  
SQ SEQUENCE 262 AA; 281.55 MW; 2D89581493BBCA30 CRC64;
```

Query Match 51.3%; Score 695.5; DB 13; Length 262;  
Best Local Similarity 58.8%; Pred. No. 3.3e-52;  
Matches 150; Conservative 23; Mismatches 51; Indels 31; Gaps 14

QY	7	HHDLSSG-----GNPPOGMPGAKMGQPADAGVPGASVGYPGQAF--PGAYGGQAF	58
Db	24	LHPQLSDLPANHPCAPPQGM-----NNPPGAPF--ATGYPATYGAAPGYG--A	74
QY	59	PGAYHGAAGAYGAPGAPGVPPGPPSGPGAYPSSGQPSABGAY--ATGGY--APAGPLTY	114
Db	75	PGPHHGPPGPYPGCG--PGYPGCG--PGYP--GCG--PGYPGCGTAPYSEAPALPLKY	127
QY	115	PYNLPPLGCGVPRMLITLTGYTKPVAANLALDFQSGNDVAFHF--PPENNNRNVAYCNK	173
QY	128	PYDLPPLGMLERLITLTGYNSPNRPSLDFKGGQDAFHFNPFEKDIHKVAYCNS	187
Db	174	LDNNMGRERQSV--FPFSGKPFKIQVLVEPDHFXYAVNDAH--LQYNHRVKKLNEISK	231
QY	188	FQNNMGKEERTAPRRPFEGTFPFKIQVLCEGDHFXYAVNDAHLLQNNPREKKLNGITKLC	247
Db	232	ISGDITLTSASYTM	246
QY	248	ISGDITLTSVLTSM	262

### RESULT 3

ID	NAME	LAST	FIRST
00000	TREMBLREL. 21,		
AC	Q8QGd9,		
01-JUN-2002	(Tremblrel. 21, Created)		
DT	01-JUN-2002 (Tremblrel. 21, last sequence update)		

D1 01-CC1-2003 (116Mb;rel. 25, last annotation update)  
 DE Galectin-3TM1 isoform containing transmembrane spanning domain.  
 CC Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

UC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;  
OC Gallus.  
OX NCBT\_TaxID=9031;  
PN [1]

RT "New alternatively spliced form of gelsolin-3, a member of beta-actin/cadherin-binding protein family contains unspliced

DR EMBL; AR479554; AAL91920.1; -; membrane; TM; RT transmembrane spanning domain and leucine zipper motif.; RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR InterPro: IPR008985; ConA\_like\_lect\_gl.  
DR InterPro: IPR001079; Galectin.  
DR Pfam: PF00337; Gal-bind\_lectin; 1.

DR PROSITE, PS00309; GALAPTIN; 1.  
KW Transmembrane.  
SEQUENCE 332 AA; 35484 MW; CE3203BCG99418CBI CRC64;  
SQ

Query Match	48.9%;	Score 663.5;	DB 13;	Length 332;
Best local Similarity	46.5%;	Pred. No. 2.6e-49;		
Matches 151; Conservative	24;	Mismatches 49;	Indels 101;	Gaps 15

QY 7 LHDALISG-----GNENPGWPGAMGNOPAGAGTYPGASYPGPGAP--PGAYPGAP 58

Db 24 LHPQSDALPAHNPGAPPPPGW----NRPPGPGAP--ATPGYFGAYPGAPGYPFG--A 74

QY 59 PGAYHGAPGAYPGADAPGVYPGPESGPGAYSSGQSDAPGAY--ATGPYG-APACPLI- 113

Db 75 PGRHNGPPGYPGGG-PGYPYGGP--GGPYR--GGP-PGYPGGGPPAPYSSAPAAVLVT 127

QY	114	-----	113
Dh	128	AtSYERGEFCESSSTCTAMGATTGCGVSIAGGMSHIAWMSNDPCTTVRPVILANVLECT	187

QY 114 -----VPYNIPLPGGVPRMITITLITGVKPNARIALDFOGRDVAHF-PREVEN 163  
168 -----PPYNIPLPGGVPRMITITLITGVKPNARIALDFOGRDVAHF-PREVEN 163  
172 -----PPYNIPLPGGVPRMITITLITGVKPNARIALDFOGRDVAHF-PREVEN 163

QY 164 NRRIVCNTKLDNNNGREEROSV-PPESGKPEFKIQLVNEPDHFKYAVNDAH-IQYNHRV 221

Db 248 HRRVIVCSNMFQNNWKEERTAPRPFEPGTEPKIQLVLCGGHFKYAVNDAHLQNPBRE 307  
 QY 222 KKLNEISKIGSDIDLTASTMTI 246  
 Db 308 KKLNEITKLCIAGDITLTSLTSMI 332

## RESULT 4

086TYS PRELIMINARY; PRT; 121 AA.  
 ID 086TYS  
 AC 086TYS;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Human full-length cDNA clone CS0D1041YE05 of placenta of Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Genoscope;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Li W.B., Gruber C., Jesse J., Polayes D.;  
 RT "Full-length cDNA libraries and normalization";  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BX161455; CAD61918.1; -  
 DR GO: GO:0046821; C:extrachromosomal DNA; IEA.  
 DR InterPro: IPR008985; Cona like lec\_g1.  
 DR InterPro: IPR010179; Galectin.  
 DR Pfam: PF00337; Gal-bind lectin.  
 DR SMART: SM00276; GLECT; 1.  
 DR PROSITE: PS00309; GALACTIN; 1.  
 KW Plasmid  
 SQ SEQUENCE 121 AA; 13897 MW; 020099E4DA6D32FA CRC64;

Query Match 44.4%; Score 603; DB 4; Length 121;  
 Best Local Similarity 98.3%; Pred. No. 1.3e-44;  
 Matches 119; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 128 MLTITIGTYKPNRIRALDFQRNDVAFHP-PRFNNRRVIVCNTKLDNNWGREEROSV 186  
 Db 1 MLTITIGTYKPNRIRALDFQRNDVAFHPNRRVIVCNTKLDNNWGREEROSV 60  
 QY 187 FPESGKPFKIQVLEPDHFKYAVNDAH-LQYHRYKCLNEISKIGSDIDLTASYTM 245  
 Db 61 FPESGKPFKIQVLEPDHFKYAVNDAHLYQYHRYKCLNEISKIGSDIDLTASYTM 120  
 QY 246 I 246  
 Db 121 I 121

## RESULT 5

081XB9 PRELIMINARY; PRT; 108 AA.  
 ID 081XB9  
 AC 081XB9;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Galectin 3 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Neugebauer S., Daimon M., Baba T., Watanabe T.;

RT "A common mutation of an AGE-binding protein is not associated with  
 RT diabetic microvascular complications in type 2 diabetes."  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB086821; BAC45249.1; -  
 FT NON-TER 1  
 FT NON-TER 1  
 SQ SEQUENCE 108 AA; 10035 MW; 5D2BE70C2CF626C5 CRC64;

Query Match 42.8%; Score 581; DB 4; Length 108;  
 Best Local Similarity 96.3%; Pred. No. 9.1e-43;  
 Matches 104; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 7 LHDALSSGNNPQGWPGWGNQAPAGAGYVPG-YPQAPPGATPGQAPPGAYHGA 65  
 Db 1 LHDALSSGNNPQGWPGWGNQAPAGAGYVPGAPPGQAPPGAYHGA 60  
 QY 66 PGAYPGAPPGVPGPSPGAYPSGQSPAPGAY-ATGPAGAPGL 112  
 Db 61 PGAYPGAPPGVPGPSPGAYPSGQSPATGAYPATGAYGAPGL 108

## RESULT 6

07ZSYL PRELIMINARY; PRT; 261 AA.  
 ID 07ZSYL  
 AC 07ZSYL;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Similar to lectin, galactose binding, soluble 3 (Galectin family  
 DE xgalectin-VIIa).  
 GN XGALECTIN-VIIA.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22552404; PubMed=12538594;  
 RX Shoji H., Nishi N., Hirashima M., Nakamura T.;  
 RA "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY  
 RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING  
 RT EMBRYOGENESIS.";  
 RL J. Biol. Chem. 278:12285-12293 (2003).  
 DR EMBL: BC046662; AAH46662.1; -  
 DR EMBL: AB080020; BAC55886.1; -  
 DR GO: GO:0005529; F:sugar binding; IEA.  
 DR GO: GO:0007157; P:heterophilic cell adhesion; IEA.  
 DR InterPro: IPR008985; Cona like lec\_g1.  
 DR InterPro: IPR010179; Galectin.  
 DR Pfam: PF00337; Gal-bind lectin.  
 DR SMART: SM00276; GLECT; 1.  
 DR PROSITE: PS00309; GALACTIN; 1.  
 KW Lectin.  
 SQ SEQUENCE 261 AA; 28332 MW; EAF78A15BE4DB91 CRC64;

Query Match 42.8%; Score 580.5; DB 13; Length 261;  
 Best Local Similarity 49.1%; Pred. No. 3e-42;  
 Matches 135; Conservative 31; Mismatches 66; Indels 43; Gaps 15;

QY 1 MADNPSLDAVS--GSGNP-----NPQGWGAWGN-QPAGAGYPGA---SYPGYRQ 47  
 Db 1 MADNPSLDAVNTGTANPQSQDPGQPPQPMAGNAPPGYPPYGAAPPGQYRYPGA 60  
 QY 48 AP-----PGAYPGAPPGAYHGAAPG-AYGADAPGV-YFG-PPSGGAYPSGQPSA 96  
 Db 61 APGGQGYVPGFPGYPGQ-----QYRPGFPGQGFPGFAPPGQGYGIGIAPQPPQKXGVPGQDEP 116

```

OY 97 PGAYATGPGAPGP--LIVPNYLPPLGGVPPRLITITIGTVKPNANRIALDFORANDVA 154
DB 117 P-----KSPAPVQLKVPYDLPPLSGVPPRLMITIGTVNPNPRFVADKRGQDIA 167
OY 155 FHF-PRFNNRRVIVYNTLTDNNNGEEROS--VFPEESGPKPKIQLVPEPDKVAVND 212
DB 168 LHNPRFDE-RPNIVNNSMIGRMWGEERKAPKPFVACQPKIQLVLCADHKKVGINN 226
OY 213 AHL-QVNRVYKLNLSKLGISGDIDLTSASYMTI 246
DB 227 ENLFQYVRYRRLHLSRVSICIGGVTLTDANVTAV 261

RESULT 7
O61357 PRELIMINARY; PRT; 139 AA.
AC Q61357;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Beta-galactoside-binding lectin (L-34) (Fragment).
GN LGLB3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88080093; PubMed=3335026;
RA Raz A., Carmi P., Pazerint G.;
RT "Expression of two different endogenous galactoside-binding lectins
sharing sequence homology."
RL Cancer Res. 48:645-649(1988).
DR EMBL: M33215; AAA37314.1; -.
DR PIR: S08576; S08576.
DR HSSP: P17931; IAAK.
DR MGD; MGI:96778; Lgal3.
DR GO; GO:0005737; Cytoplasm; IDA.
DR GO; GO:0005515; F:protein binding; IPT.
DR InterPro; IPR008985; Cona_like_lect_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 1.
DR SMART; SM00276; GLECT; 1.
DR PROSITE; PS00309; GALAPTIN; 1.
KW Lectin.
KW NON TER.
SQ SEQUENCE 139 AA; 15914 MW; B6373F3DC378568 CRC64;

Query Match 32.7%; Score 444; DB 11; Length 139;
Best Local Similarity 76.8%; Pred. No. 9e-31; Indels 2; Gaps 2;
Matches 86; Conservative 7; Mismatches 17;

OY 104 PGAPAGPLIVPNYLPPLGGVPPRLITITIGTVKPNANRIALDFORANDVAHF-PRFNE 162
DB 22 PIMCPMTTVDYDLPPLSGVPPRLITITIGTVKPNANRIALDFORANDVAHFPRFNE 81

OY 163 NNRRVIVYNTLTDNNNGEEROSVFPEESGPKPKIQLVPEPDKVAVND 214
DB 82 NNRRVIVYNTLTDNNNGEEROSVFPEESGPKPKIQLVLAADHSGCG-HDAH 132

RESULT 8
O99T11 PRELIMINARY; PRT; 341 AA.
AC O99T11;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Galectin like protein.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

```

```

OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Head kidney;
RA Soma G., Inagawa H., Honda T., Nakanishi T., Otake M., Nishizawa T.,
RA Kanou J., Endo M.;
RT "Rainbow trout CDNA similar to galectin 9 like."
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB027452; BAA8670.1; -.
DR HSSP; P47929; IBAK.
DR InterPro; IPR008985; Cona_like_lect_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 341 AA; 37816 MW; 176D8358C1950C4 CRC64;

Query Match 23.2%; Score 314.5; DB 13; Length 341;
Best Local Similarity 36.0%; Pred. No. 4.4e-19;
Matches 90; Conservative 30; Mismatches 69; Indels 61; Gaps 17;

OY 9 DALSGG-----NPN-----QWPGAMGQPGAGVPGASYPGPGAPGAY 53
DB 132 DTISADSKVLTSLVFONPAPTTPAQGFPA-----QF-----GFP--SYGFPAG--RG-F 178
OY 54 PGQAPGAYHGAAPGAPGAPGVPYPPGPGGAYPSSGQPSAPGAYATGPY-GAPAGPL 112
DB 179 P-----SYGFPFA--QGFPPSCPG-FP--GGPFG-----PYGFPAGP- 211
OY 113 TVPNYLPPLGGVPPRLITITIGTVKPNANRIALDFORANDVAHF-PRFNNRRVIVCN 171
DB 212 AVPYKNMIGLTVPGRIINIGVNPANRRPHINLPSGIALHNPREDT--LVARN 268
OY 172 TKLNNWGERERQGVPPESGPKPKIQLVPEPDKVAVNDHQLQ-YNHRVYKLNLSKLG 230
DB 269 SKLPDQWKEKESGGMPPHSGAFTLSTCDACQCYKIVYNNQNSIYHRRITLLQVVYL 328
OY 231 GISGDIDLTS 240
DB 329 EVDGLSLTS 338

RESULT 9
O99L83 PRELIMINARY; PRT; 322 AA.
AC O99L83;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Lectin, galactose binding, soluble 9.
GN LgalS9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strusberg R.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC003754; AAH03754.1; -.
DR HSSP; P17931; IAAK.
DR MGD; MGI:109496; LgalS9.
DR InterPro; IPR008985; Cona_like_lect_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
SQ SEQUENCE 322 AA; 36545 MW; 032D77400737562E CRC64;

Query Match 22.0%; Score 299; DB 11; Length 322;
Best Local Similarity 37.5%; Pred. No. 8.9e-18;
Matches 69; Conservative 28; Mismatches 47; Indels 40; Gaps 8;

```



Db 292 NDIDILEICGDLQTS 307

# RESULT 13

Q8WYQ7 08WYQ7 PRELIMINARY; PRT; 311 AA.  
AC 08WYQ7;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Galectin-9.  
GN LGALS9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RT Homo sapiens galectin-9 (LGALS9) / ecalectin gene, exon 2 through  
RT 11."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB040130; BAB83623.1; -.  
DR EMBL; AB040129; BAB83623.1; JOINED.  
DR InterPro; IPR008985; Cona\_like\_1ec\_g1.  
DR InterPro; IPR001079; Galectin.  
DR Pfam; PF00337; Gal-bind\_lectin; 2.  
DR SMART; SM00276; GLECT; 2.  
DR PROSITE; PS00309; GALAPTIN; 2.  
SQ SEQUENCE 311 AA; 34690 MW; 3D00232FE39D776C CRC64;

Query Match 19.5%; Score 264; DB 4; Length 311;  
Best Local Similarity 36.5%; Pred. No. 9, 1e-15;  
Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;  
QY 65 ARGATPGAPA-PGYVPPGSPGAYPSSGQSPAGAVATGYGAPAGPLVPPNLPFG 122  
Db 159 AFGQWFSTPAIPPMYPHP-----AP-----MPFITILIG 189  
QY 123 GVPPRMILITLGTGVNPNARIALDFQGNDAVHF-FRENNRRVIVNTKLDNNWGRE 181  
Db 190 GLYPSKILSLSTVLPASARFHINLCSGNHIAFHINPRDEN--AVVNTQIDNSWGSSE 246  
QY 182 ERQ--SVPEFESGKPEKIOVLVEPDHFKVAVNDAH-QYNHRVKKLINEISKIGISDIDL 238  
Db 247 EESLPRKMFVGGQSPFWILCEAHGLKVAVDGQHLFEYVYHRLNLPITNLEVGGLDIL 306  
QY 239 T 239  
Db 307 T 307

# RESULT 14

Q9XSM9 09XSM9 PRELIMINARY; PRT; 349 AA.  
AC 09XSM9;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Urate transporter/channel protein, isoform (UATp.1).  
GN UATP.1.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
OX NCBI\_TaxID=9623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RT Urate transporter/channel protein, isoform (UATp.1).  
RT transporter/channel from cultured LLC-PK1 epithelial cells."  
RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ131827; CAB44279.1; -.

DR HSSP; P17931; 1A3X.  
DR InterPro; IPR008985; Cona\_like\_1ec\_g1.  
DR InterPro; IPR001079; Galectin.  
DR Pfam; PF00337; Gal-bind\_lectin; 2.  
DR SMART; SM00276; GLECT; 2.  
DR PROSITE; PS00309; GALAPTIN; 1.  
SQ SEQUENCE 349 AA; 38899 MW; BF83D3E213E7B64C CRC64;

Query Match 19.2%; Score 260; DB 6; Length 349;  
Best Local Similarity 31.7%; Pred. No. 2, 3e-14;  
Matches 66; Conservative 32; Mismatches 68; Indels 42; Gaps 9;

QY 37 PGASYPGPGQ--APPGAYEGQAPP-----GAYHGAAGAYGAPAGVYPPSGPGAY 88  
Db 167 PACPPPHKGRKRPGRWRPANSAPITQVITHVOSTPGQ-----MFENPMIPWAY 218  
QY 89 PSSGQSPAGAVATGYGAPAGPLIVPPNLPFGVPPRMILITLGTVPNANRIALDFQ 148  
Db 219 FN-----PVFP--IPFFASIPGGLPSPKSIWSTGLPSPQSTFYINLR 259  
QY 149 RGNDAVHF-FRENNRRVIVNTKLDNNWGREER--QSVPEFESGKPEKIOVLVEPDH 205  
Db 260 GSDIAFHLNPRFEN--AVVNTQIDSSWGPBERGLPRKMPFRRGQSLWLILCESHC 316  
QY 206 FKAVANDAH-QYNHRVKKLINEISKIGI 232  
Db 317 FKAVDQGHLEFYVYHRLKHLPTINSLEV 344

# RESULT 15

Q7ZTB5 07ZTB5 PRELIMINARY; PRT; 315 AA.  
AC 07ZTB5;  
DT 01-JUN-2003 (TReMBLrel. 24, Created)  
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Galectin family xgalectin-VIIIA.  
GN XGALECTIN-VIIIA.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-22552404; PubMed-12538594;  
RA Shoji H., Nishi N., Hiraishima M., Nakamura T.;  
RT "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY  
RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING  
RT EMBRYOGENESIS.";  
RT J. Biol. Chem. 278:12285-12293 (2003).  
DR EMBL; AB080021; BAC55887.1; -.  
DR InterPro; IPR008985; Cona\_like\_1ec\_g1.  
DR InterPro; IPR001079; Galectin.  
DR Pfam; PF00337; Gal-bind\_lectin; 2.  
DR SMART; SM00276; GLECT; 2.  
SQ SEQUENCE 315 AA; 35895 MW; 3A42275840579A9A CRC64;

Query Match 19.0%; Score 258.5; DB 13; Length 315;  
Best Local Similarity 43.2%; Pred. No. 2, 8e-14;  
Matches 60; Conservative 21; Mismatches 47; Indels 11; Gaps 5;

QY 113 IVPYNIPLPGVPPRMILITLGTVPNANRIALDFQGN-----DVAFHF-FRENNNR 165  
Db 14 VEVYVGTIGLELPGMIVIHGVHDADRFQDFQGNQSVQPSDVAFHFNFPGSGSH 73  
QY 166 RVIVNPKLDNNWGREEROSVPEFESGKPEKIOVLVEPDHFKVAVNDAH-QYNHRVKK 224  
Db 74 --IVCNTLENEKKGWKEKTYQMPFTYGOPELILFVFDHDFQVSSNGKMLLVYKRI-SL 130  
QY 225 NEISKIGISDIDLTLASY 243  
Db 131 QVVDYTLGISGKVKINTIGF 149

RESULT 16	Q8UW98	PRELIMINARY;	PRT;	343 AA.
ID	Q8UW98			
AC	Q8UW98;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Galectin family xgalectin-IIIA.			
GN	XGALECTIN-IIIA.			
OS	Xenopus laevis (African clawed frog).			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;			
OC	Xenopodinae, Xenopus.			
OX	NCBI_Taxid=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RA	Shoji H., Nishi N., Hirashima M., Nakamura T.;			
RT	"Purification and cDNA cloning of Xenopus liver galectins, and their expression in adult tissues.";			
RL	Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AB060971; BAB83258.1; -			
DR	InterPro; IPR008985; Consa like lec_g1.			
DR	InterPro; IPR01079; Galectin.			
DR	Pfam; PF00337; Gal-blind lectin; 2.			
DR	SMART; SM00276; GLECT; 2.			
SO	SEQUENCE 343 AA; 38361 MW; B55056ED8F063589 CRC64;			

```

Db      258 LITENS---VYVNSPLGSGWGEERQLSYNPPRPGQYFDISLRGMDRFTYFVNGQFCDF 314
QY      218 NHRVYKLNIEISKLGISGDI 236
Db      315 AHRYSMLQMDIRIBENGV 333

RESULT 19
Q91X74      PRELIMINARY;      PRT;      326 AA.
ID      Q91X74
AC      Q91X74
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Lectin, galactose binding, soluble 4 (Galectin-4) (Hypothetical
      protein).
GN      LGALS4.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Colon;
RA      Strausberg R.;
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RT      [2]
RM      SEQUENCE FROM N.A.
RA      STRAIN=BALB/c; TISSUE=Liver;
      Maly P., Jenikova G., Cummings R.D.;
      "Molecular cloning and tissue distribution of mouse galectin-4.";
      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC011236; AAH11236.1; -
DR      EMBL; BC021632; AAH21632.1; -
DR      EMBL; BC030297; AAH30297.1; -
DR      EMBL; AY044870; AAK97790.1; -
DR      MGD; MGI:107536; Lgals4.
DR      InterPro; IPR008985; Cona_like_1ec_g1.
DR      InterPro; IPR01079; Galectin.
DR      Pfam; PF00337; Gal-bind_lectin; 2.
DR      PROSITE; PS00309; GALAPTIN; 2.
KW      Hypothetical protein
SQ      SEQUENCE      326 AA; 36372 MW; 7F3DD89862A851B5 CRC64;

Query Match      17.9%; Score 243.5; DB 11; Length 326;
Best Local Similarity 35.8%; Pred. No. 5.8e-13;
Matches 54; Conservative 31; Mismatches 57; Indels 9; Gaps 5;

QY      99 AYATGPGAGPLIVPNLPPLPGGVVPRMLITLIGTVENARIALDF-----QRGNDVA 154
Db      2 AYVPAAGYQFTYNPPLPYKRPFGLSVGSVYIQGAKENMKRFHVNPAVSGDDADVA 61
QY      155 FHF-PPFENNRVIVCNTKLDNNMGREERQSVFPESGKPFKIQVLVEPDHFKVAVN-D 212
Db      62 FHNPPFPDGDK--VFNTMQSGQMGKEKKKMPFGKGFELVEMVMEHHKKVYVNGN 119
QY      213 AHIQYNRVKLNIEISKLGISGDIIDLTASY 243
Db      120 SFYEYGHRL-PYQWVTHLQVDGLLEQSLNF 149

RESULT 20
Q7ZTB9      PRELIMINARY;      PRT;      328 AA.
ID      Q7ZTB9
AC      Q7ZTB9
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Galectin family xgalectin-11b.
GN      XGALLECTIN-11b.
OS      Xenopus laevis (African clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC      Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
      Xenopodinae; Xenopus.
CX      NCBI_TaxId=8355;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22552404; PubMed=12538594;
      Shoji H., Nishi N., Hirashima M., Nakamura T.;
      "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY
      DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING
      EMBRYOGENESIS.";
      RT      J. Biol. Chem. 278:12285-12293 (2003).
      RL      U. Biol. Chem. 278:12285-12293 (2003).
      DR      EMBL; AB080016; BACS5882.1; -
      DR      InterPro; IPR008985; Cona_like_1ec_g1.
      DR      InterPro; IPR01079; Galectin.
      DR      Pfam; PF00337; Gal-bind_lectin; 2.
      DR      SMART; SM00276; GLECT; 2.
SQ      SEQUENCE      328 AA; 36194 MW; F70DF7556ABEFA0 CRC64;

Query Match      17.6%; Score 238.5; DB 13; Length 328;
Best Local Similarity 28.8%; Pred. No. 1.6e-12;
Matches 75; Conservative 33; Mismatches 91; Indels 61; Gaps 11;

QY      24 GAWGNQPRAGAGYP-----GASYPGYQGAP-----PG 51
Db      82 GSWGSEERKXKDSFPNKKKSFELTMINSSFEITVNGSSFYKFRHRLPLERVDSLOING 141
QY      52 AYPGAPPGAYHGAAPGAYGAP-AP-----GYYPGPGSGAYPSSQPSAPGAYATGP 104
Db      142 DVTYQSLTIAGGGGGGMAAPMAFMFPNGVWMLPVPFPAAGNLFVWGPF-----V 192
QY      105 YGAPAGPLIVPNLPPLPGGVVPRMLITLIGTVKPRANIALDFQSG--NDVAFHF-PPFN 161
Db      193 YNPP-----VPYNGTIOGLPFRKTVIRGFLPEGAEFHNFKAGSSNDLALHNPRLT 247
QY      162 ENNRVIVICNTKLDNNMGREERQ-SVFPESGKPFKIQVLVEPDHFKVAVNDAHL-QYNN 219
Db      248 EN--AVVNSQLGHWGGEERQLSYNPPRAGQYDILRGMDRFTYFVNGQFCDFAH 304
QY      220 RVKLNIEISKLGISGDIIDLT 239
Db      305 RYSMQMDIRLEVEGNVVLVS 324

RESULT 21
Q8K419      PRELIMINARY;      PRT;      326 AA.
ID      Q8K419
AC      Q8K419
DT      01-OCT-2002 (TrEMBLrel. 22, Created)
DT      01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Galectin-4.
GN      LGALS4.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6; TISSUE=Colon;
      Hokama A., Taraka Y., Mizoguchi A.;
      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF510729; AAM44060.1; -
DR      MGD; MGI:107536; Lgals4.
DR      InterPro; IPR008985; Cona_like_1ec_g1.
DR      InterPro; IPR01079; Galectin.
DR      Pfam; PF00337; Gal-bind_lectin; 2.
DR      SMART; SM00276; GLECT; 2.
DR      PROSITE; PS00309; GALAPTIN; 2.
SQ      SEQUENCE      326 AA; 36386 MW; 7F3C899862AA23B5 CRC64;

Query Match      17.5%; Score 237.5; DB 11; Length 326;
Best Local Similarity 35.1%; Pred. No. 1.9e-12;
Matches 53; Conservative 31; Mismatches 58; Indels 9; Gaps 5;

```



```

QY 99 AATGTYGAPAGPLIVPNILPFGVVPRLITITGTVKPNANRIALDF----QKGNDA 154
DB 2 AAVPAGGYOPTYNPLTPYKRPILPGSLVGSVSVYIQMAKENMRPFVNFVAVGDDGADVA 61
QY 155 FHF-PFENNRRVIVCNTKLDNNWGEROSVFPFESGKPFKIQVLVEPDHFKVAVN-D 212
DB 62 FHFNPFDGWDK--VFETKMGSGQWKEKKKMPQKGFELVPMVPMBEHKKVAVNGN 119
QY 213 AHLQYNHRVYKLNLSKLGISGDIIDLTASY 243
DB 120 SFPEYGHRL-PVQMTLHLQVDDDELQSLNF 149

RESULT 22
Q9TUB8 PRELIMINARY; PRT; 328 AA.
AC Q9TUB8
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Galectin-4.
GN LGALS4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New England White;
RX MEDLINE=20147368; PubMed=10683764;
RA Jiang W., Puch S., Guo X., Bhavanandan V.P.;
RT "Signature sequences for the galectin-4 subfamily.";
RL IUBMB Life 48:601-605(1999).
DR HSSB; P47929; IRLKZ.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
SQ SEQUENCE 328 AA; 36462 MW; 030BF1B723D2ACF CRC64;

Query Match 17.2%; Score 233.5; DB 6; Length 328;
Best Local Similarity 34.5%; Pred. No. 4.3e-12;
Matches 68; Conservative 23; Mismatches 81; Indels 25; Gaps 10;

QY 55 GOAPGAYHAGAPGAPGAPVYRPPSGPAGYSSQPSAPGAYATGPGAPAGPLIV 114
DB 152 GQPTP--HQRPTMGSPSPSY--GHP-GYGAQGLHSLPSMEGPTPNP-----PV 197
QY 115 PYNLEPLPGVVPRLITITGTVKPNANRIALDFQGN--DVAFHF-PFENNRRVIVCN 171
DB 198 PFTGLQGGILARLITIKGVYPRIGKSFILNFKVSGSDLAHINPMTEG--VVVRN 254
QY 172 TKLDNNWGEROSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNHRVYKLNLSK 229
DB 255 SLLGSSWGAERKKAAYNPFQGYFDLSIRGMDRFKYANGQLFDYAHFFPAQKVDV 314
QY 230 LGISGDIIDLTASYTMI 246
DB 315 IEIQGDVAL--SVQI 328

RESULT 23
Q9TUB8 PRELIMINARY; PRT; 308 AA.
AC Q9TUB8
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Galectin family xgalectin-11b.
GN XGALECTIN-11b.

```

```

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxId=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=2252404; PubMed=12538594;
RA Shoji H., Nishi N., Hirashima M., Nakamura T.;
RT "Characterization of the Xenopus Galectin Family. THREE STRUCTURALLY
RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING
RT EMBRYOGENESIS.";
RL J. Biol. Chem. 278:12285-12293 (2003).
DR EMBL; AB080017; BAC55883.1; -.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
SQ SEQUENCE 308 AA; 34808 MW; A1C2D04079048AE CRC64;

Query Match 16.9%; Score 230; DB 13; Length 308;
Best Local Similarity 37.6%; Pred. No. 7.9e-12;
Matches 53; Conservative 30; Mismatches 48; Indels 10; Gaps 6;

QY 105 YGAPAGPLIVPNILPFGVVPRLITITGTVKPNANRIALDFQGN--NDVAFHF-PR 159
DB 3 FGAINNP-PVPTPLPLSHGLHDGFLVYSGAVTHSGDFNFNFQCHGSNDIAHFHNR 61
QY 160 FENNRRVIVCNTKLDNNWGEROSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYN 218
DB 62 FIDGG--IVVCNTERQSGWKEKREMPFHHQPEFELRLVTHNSYVSVVRNHYLEVH 119
QY 219 HRVYKLNLSKLGISGDIIDLT 239
DB 120 HRI-PIQKVNMTLIGCVALT 139

RESULT 24
Q9TEV1 PRELIMINARY; PRT; 359 AA.
AC Q9TEV1
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Lymphocyte/NHL galectin-8 long isoform.
GN LGALS8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Moisan S., Mercier J., Demers M., Belanger S.D., Alain T.,
RA Kosciakowska A.E., Potworowski E.F., St-Pierre Y.;
RT "Galectins in murine and human non-Hodgkin's lymphomas.";
RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF468213; AAL7076.1; -.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 359 AA; 40151 MW; F87EBFE92E46F571 CRC64;

Query Match 16.8%; Score 228.5; DB 4; Length 359;
Best Local Similarity 40.0%; Pred. No. 1.3e-11;
Matches 56; Conservative 22; Mismatches 51; Indels 11; Gaps 5;

QY 113 IVPYNLPGGVVPRLITITGTVKPNANRIALDFQGN-----DVAFHF-PFENNRR 165
DB 16 VIVVYVTPQULDPGTVITVCGHVPSPADDFQVLDQNGSSVKPRADVAFFHNPFKAG- 74
QY 166 RVIVCNTKLDNNWGEROSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNHRVYK 224

```

```

Db      75 -CIVCNLTLINEKMGREBEITYDTPFKREKSFELIVMWLKDQFOVAVNGKATLLYGRIGP- 132
      225 NEISKLGISGIDLTLSASYT 244
      133 EKIDTLGIYGVKNVHISIGFS 152

RESULT 25
Q9BXC8 PRELIMINARY; PRT; 359 AA.
AC Q9BXC8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Colorectal carcinoma-derived galectin-8 variant II.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA Lahm H., Siebert H.-C., Andre S., Hoeflich A., Diehl D., Sordat B.,
RA Kallner H., Wolf E., Gabius H.-J.;
RT "Coca (colorectal carcinoma-derived) galectin-8 variant II.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF342816; AAK1636.1; -.
DR HSSP; P17931; IAKK.
DR InterPro; IPR008985; ConA like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-blind lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 359 AA; 40363 MW; 39BBD61411FAD98 CRC64;

Query Match 16.5%; Score 223.5; DB 4; Length 359;
Best Local Similarity 39.3%; Pred. No. 3.5e-11;
Matches 55; Conservative 23; Mismatches 51; Indels 11; Gaps 5;

QY 113 IVEYNLPDGGVVRMLITLLGTGVKNANRISALDPQRG-----DVAFHF-FRENNNR 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 VIFPVGITIPQDLPGLTIVRGHVPSDADFQVDLQNGSSMKPRADVAEFHFNRFPRAG- 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 166 RVLVCTKLDNMNGRRRQSVFPFESGKPKIQVLEPDPFKAVNDAH-LQYNNRVKKL 224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 -CIVCNLTLINEKMGREBEITYDTPFKREKSFELIVMWLKDQFOVAVNGKATLLYGRIGP- 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 225 NEISKLGISGIDLTLSASYT 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 EKIDTLGIYGVKNVHISIGFS 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 26
Q7SYPO PRELIMINARY; PRT; 204 AA.
AC Q7SYPO;
DT 01-OCT-2003 (TEMBLrel. 25, Created)
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Klein S.U., Strussberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus

```

```

RT intrative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX MEDLINE=22188257; PubMed=12477932;
RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buehler K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarnatone P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalon D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy U., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maitra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Klein S., Strussberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054324; AAH54324.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 204 AA; 22637 MW; 82532249A1339BE1 CRC64;

Query Match 16.2%; Score 219.5; DB 13; Length 204;
Best Local Similarity 33.3%; Pred. No. 3.9e-11;
Matches 57; Conservative 25; Mismatches 74; Indels 15; Gaps 7;

QY 73 PARGVPRPSPRGAVPSSGQSPAPGAVATGYGAPAGLIVPNLPLPGVVRMLITL 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 39 PASGA--GAKAKAGLIMSSTOENLPW----LGPILHPLIPFKAMITGMIIPKKTVM 91
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 LGTVKPNANRISALDPQRG--NDVAFHF-FRENNNRVIVCTKLDNMNGRRRQSV-FE 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 KGIYVNNAKNFOISFRVGTINDALHINPLKN--TLIRNSFINGWGEKQVYKNP 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 FESGKPEFKIQVLEPDPFKAVNDAH-LQYNNRVKKLNEISKLGISGIDLT 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 FHQGEHFDISIRSGEKQYKVVYVNGVCFVYPHRLTNLQGVDTLEADGDIKL 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 27
Q7ZTB6 PRELIMINARY; PRT; 319 AA.
AC Q7ZTB6;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Galectin family xgalectin-VIIA.
GN XGALACTIN-VIIA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22552404; PubMed=12538594;
RA Shoji H., Nishi N., Hiraehima M., Nakamura T.;
RT "Characterization of the Xenopus galectin family: THREE STRUCTURALLY
RT DIFFERENT TYPES AS IN MAMMALS AND REGULATED EXPRESSION DURING

```



DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Galectin family xgalectin-IVA.  
 GN XGALECTIN-IVA.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Xenopodidae; Xenopus.  
 OC NCBI\_TaxID=8355;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC Tissue: Liver;  
 RA Shoji H., Nishi N., Hirashima M., Nakamura T.;  
 RT "Purification and cDNA cloning of Xenopus liver galectins, and their  
 RT expression in adult tissues."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB060972; BAB83259.1; -  
 DR InterPro; IPR008985; Consilium\_1ec\_g1.  
 DR InterPro; IPR001079; Galectin.  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.  
 DR SMART; SM00276; GLECT; 2.  
 SQ SEQUENCE 332 AA; 36822 MW; D273FD35484B9C2 CRC64;

Query Match 15.7%; Score 213; DB 13; Length 332;  
 Best Local Similarity 37.3%; Pred. No. 2.6e-10;  
 Matches 50; Conservative 25; Mismatches 53; Indels 6; Gaps 5;

QY 114 VYNNLPFGVVRMLITLITGVKPNANRIALDFQ-RGNDVAFH-FRENNNRVYVGN 171  
 AC Q83467; PRELIMINARY; PRT; 703 AA.  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Porcine adenovirus 4 putative fiber protein.  
 OS Porcine adenovirus 4.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OC NCBI\_TaxID=35267;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PAV-4;  
 RA MEDLINE=96435006; PubMed=8837892;  
 RA Kleiboecker S.B.;  
 RT "Sequence analysis of the fiber genomic region of a porcine adenovirus  
 RT predicts a novel fiber protein."  
 RL Virus Res. 39:299-309(1995).  
 DR EMBL; U25120; AAB40001.1; -  
 DR HSSP; P17931; 1A3K.  
 DR GO; GO:0008037; P-cell recognition; IEA.  
 DR InterPro; IPR000939; Adeno\_fiber2.  
 DR InterPro; IPR000978; Adeno\_fiber\_knob.  
 DR InterPro; IPR000931; Adeno\_fiber.  
 DR InterPro; IPR008985; Consilium\_1ec\_g1.  
 DR InterPro; IPR001079; Galectin.  
 DR InterPro; IPR008982; Viral\_act.  
 DR InterPro; IPR009013; Viral\_act\_shift.  
 DR Pfam; PF00541; adeno\_fiber; 1.  
 DR Pfam; PF00608; adeno\_fiber2; 1.  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.

RESULT 31

Q83467; PRELIMINARY; PRT; 703 AA.  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Porcine adenovirus 4 putative fiber protein.  
 OS Porcine adenovirus 4.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OC NCBI\_TaxID=35267;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PAV-4;  
 RA MEDLINE=96435006; PubMed=8837892;  
 RA Kleiboecker S.B.;  
 RT "Sequence analysis of the fiber genomic region of a porcine adenovirus  
 RT predicts a novel fiber protein."  
 RL Virus Res. 39:299-309(1995).  
 DR EMBL; U25120; AAB40001.1; -  
 DR HSSP; P17931; 1A3K.  
 DR GO; GO:0008037; P-cell recognition; IEA.  
 DR InterPro; IPR000939; Adeno\_fiber2.  
 DR InterPro; IPR000978; Adeno\_fiber\_knob.  
 DR InterPro; IPR000931; Adeno\_fiber.  
 DR InterPro; IPR008985; Consilium\_1ec\_g1.  
 DR InterPro; IPR001079; Galectin.  
 DR InterPro; IPR008982; Viral\_act.  
 DR InterPro; IPR009013; Viral\_act\_shift.  
 DR Pfam; PF00541; adeno\_fiber; 1.  
 DR Pfam; PF00608; adeno\_fiber2; 1.  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.

DR PRINTS; PF00307; ADENOVIFIBRE.  
 DR SMART; SM00276; GLECT; 2.  
 SQ SEQUENCE 703 AA; 76649 MW; C2D9E7A8AEB3C0E1 CRC64;  
 Query Match 15.5%; Score 210.5; DB 12; Length 703;  
 Best Local Similarity 32.0%; Pred. No. 1.1e-09;  
 Matches 71; Conservative 27; Mismatches 101; Indels 23; Gaps 9;

QY 37 PGASVPGGQAPRPAYGQAPRGVHGAPGA-YPGAPVGYPPPSG-----PGAYPS 90  
 DB 301 PLASAGYFGKLAASSEMPAPPEAQODQAABEPAPAEAPAEAPAEAPAEAPPRKP 360  
 QY 91 SGQSPAPAY-----ATGPGAPAGPLIV-----PYNNLPFGVVRMLITLITGVKPN 140  
 DB 361 RGDLAA--LYNRVHSDTAADPTSPPELVTLPPDFVLPFGVPTGASIVLGLTLPSA 418  
 QY 141 NRALDFQRG-NDVAFH-FRENNNRVYVGN-TDNNGREROSVFPESGKPEK 197  
 DB 419 VFTLIDLVTGPASLALHFNVRPLLEGKHIYCNSSREGSSNWGEVYRPPFPREKRPVL 478  
 QY 198 QVLEVPDHFKVAVNDAHLYNNHRYKKINEISLIGSDIDL 239  
 DB 479 VIVQSDTYQITVNGKPLV--DPQRLQGITRASLSGDLVFT 518

RESULT 32

Q8KN59 PRELIMINARY; PRT; 162 AA.

AC Q8KN59;  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Ovis aries (Sheep).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OC NCBI\_TaxID=9940;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Dunphy J.L., Barham G.J., Bischof R.J., Nash A., Meusen E.N.T.;  
 RT "Isolation and characterization of a novel eosinophil-specific  
 RT galectin."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF443208; AAL37895.1; -  
 DR InterPro; IPR008985; Consilium\_1ec\_g1.  
 DR InterPro; IPR001079; Galectin.  
 DR Pfam; PF00337; Gal-bind\_lectin; 1.  
 DR SMART; SM00276; GLECT; 1.  
 DR PROSITE; PS00309; GALACTIN; 1.  
 SQ SEQUENCE 162 AA; 18194 MW; 0372ELF0CD5CFB4C CRC64;

Query Match 15.2%; Score 206.5; DB 6; Length 162;  
 Best Local Similarity 37.1%; Pred. No. 3.9e-10;  
 Matches 53; Conservative 25; Mismatches 48; Indels 17; Gaps 8;

QY 115 PY-NLPLP-----GGVPRMLITLITGV-KENANRIALDFQRG-----NDVAFH-FREN 161  
 DB 21 PYNNLPFGVVRMLITLITGVKPNANRIALDFQ-RGNDVAFH-FRENNNRVYVGN 171  
 QY 162 ENNRVAVCNKLDNNWGREEROSVFPESGKPEKQVLEVPDHFKVAVNDA-HLYNNR 220  
 DB 81 GSG--YVVCNTMQLGHWGPEKMKQPPFGKSLFEICFYVDSSESFYVNGSIFLDYAR 138  
 QY 221 VKKLNISKLIGSGDILDSASY 243  
 DB 139 L-PFEQVNAISIGCVHVSYSIF 160

RESULT 33

Q82HF3 PRELIMINARY; PRT; 903 AA.



Q8CCV9  
ID Q8CCV9 PRELIMINARY; PRT; 415 AA.  
AC Q8CCV9;  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, last annotation update)  
DE Annexin A7.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RA MEDLINE=22354683; PubMed=12466851;  
RX The FANTOM Consortium,  
RT the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK032013; BAC27647.1; -  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IEA.  
DR InterPro: IPR001464; Annexin.  
DR Pfam; PF00191; annexin; 3.  
DR PRINTS; PR00196; ANNEXIN.  
DR ProDom; PD000143; Annexin; 3.  
DR SMART; SM00335; ANX; 3.  
DR PROSITE; PS00223; ANNEXIN; 3.  
SQ SEQUENCE 415 AA; 44468 MW; 2A61242FD6222EC6 CRC64;

Query Match 14.2%; Score 192.5; DB 11; Length 415;  
Best Local Similarity 33.0%; Pred. No. 2e-08;  
Matches 74; Conservative 16; Mismatches 77; Indels 57; Gaps 14;

QY 11 LSGSGNPNQGMFGMAGNQPAG-----AGGPGASVPGYPGQAP--GAYPGQAPPG 60  
DB 1 MSYPPYP-PTGYPPPPGYPAGQESSFPTAGQYP-----YSGPFPMGAGVP-PAPSG 52  
QY 61 AYHGAPGAVP---GAPAPGVYPG--PSPGPGAYPS-----SGQPSAPG-AYA 101  
DB 53 GYPGA-GGYPAPGVPAPGYPGALSPGGPAPVPGGQGFAPGAGFGSPQPPAQSYG 111  
QY 102 TGPY-----GAPAGPLVPPNPLPFGVVP-----RMLTTLTGTVKNARIALDPFR 149  
DB 112 GGPACVVPVPGFPGGQMPGQY---PGGQAPYPSQAPASMTGQTGTLIPASN---FDAMR 164  
QY 150 GNDVAFHFPPFENNRRVIVCTKLDNNMGREROSVFPFESGK 193  
DB 165 DAELIRKAMKGFDTGQALIVDVVSNRSNDQGRQIRAKFKTWGK 208

RESULT 37  
Q922A2 PRELIMINARY; PRT; 463 AA.  
ID Q922A2  
AC Q922A2;  
DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)  
DE Annexin A7.  
GN ANXA7.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBD databases.  
CC -1- DOMAIN: A PAIR OF ANNEXIN REPEATS MAY FORM ONE BINDING SITE FOR  
CC CALCIUM AND PHOSPHOLIPID (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.

DR EMBL; BC008997; AAH08997.1; -  
DR PIR; S29170; S29170.  
DR MGI; MGI:88031; Anxa7.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IEA.  
DR InterPro: IPR001464; Annexin.  
DR Pfam; PF00191; annexin; 4.  
DR PRINTS; PR00196; ANNEXIN.  
DR ProDom; PD000143; Annexin; 4.  
DR SMART; SM00335; ANX; 4.  
DR PROSITE; PS00223; ANNEXIN; 4.  
KW Annexin; Calcium-binding; Calcium/phospholipid-binding;  
KW Repeat.  
SQ SEQUENCE 463 AA; 49909 MW; D2D5FF652E85F6ED CRC64;

Query Match 14.2%; Score 192.5; DB 11; Length 463;  
Best Local Similarity 33.0%; Pred. No. 2.3e-08;  
Matches 74; Conservative 16; Mismatches 77; Indels 57; Gaps 14;

QY 11 LSGSGNPNQGMFGMAGNQPAG-----AGGPGASVPGYPGQAP--GAYPGQAPPG 60  
DB 1 MSYPPYP-PTGYPPPPGYPAGQESSFPTAGQYP-----YSGPFPMGAGVP-PAPSG 52  
QY 61 AYHGAPGAVP---GAPAPGVYPG--PSPGPGAYPS-----SGQPSAPG-AYA 101  
DB 53 GYPGA-GGYPAPGVPAPGYPGALSPGGPAPVPGGQGFAPGAGFGSPQPPAQSYG 111  
QY 102 TGPY-----GAPAGPLVPPNPLPFGVVP-----RMLTTLTGTVKNARIALDPFR 149  
DB 112 GGPACVVPVPGFPGGQMPGQY---PGGQAPYPSQAPASMTGQTGTLIPASN---FDAMR 164  
QY 150 GNDVAFHFPPFENNRRVIVCTKLDNNMGREROSVFPFESGK 193  
DB 165 DAELIRKAMKGFDTGQALIVDVVSNRSNDQGRQIRAKFKTWGK 208

RESULT 38  
Q8BP75 PRELIMINARY; PRT; 463 AA.  
ID Q8BP75  
AC Q8BP75;  
DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
DT 01-MAR-2003 (T-EMBLrel. 23, last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)  
DE Annexin A7.  
GN ANXA7.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Body;  
RA MEDLINE=22354683; PubMed=12466851;  
RX The FANTOM Consortium,  
RT the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK077579; BAC36874.1; -  
DR MGI; MGI:88031; Anxa7.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IEA.  
DR InterPro: IPR001464; Annexin.  
DR Pfam; PF00191; annexin; 4.  
DR PRINTS; PR00196; ANNEXIN.  
DR ProDom; PD000143; Annexin; 4.  
DR SMART; SM00335; ANX; 4.  
DR PROSITE; PS00223; ANNEXIN; 4.  
SQ SEQUENCE 463 AA; 49953 MW; 94A22F620AF08C50 CRC64;

Query Match 14.2%; Score 192.5; DB 11; Length 463;  
Best Local Similarity 33.0%; Pred. No. 2.3e-08;  
Matches 74; Conservative 16; Mismatches 77; Indels 57; Gaps 14;

```

QY 11 LSGSGNPNQGWPGMAGMNQAG-----AGCYPGASVYGYPGQAPP---GAYPGQAPPG 60
DB 1 MSYPPGVP-PTGYPPFPFGYPAGGESSPTTAGQYP-----YPSGFPFPMGGAPV-PAPSG 52
QY 61 AYHGAGGAVP---GAPAGVYYPG--PPSGGAPYS-----SGQSPAPG-AYA 101
DB 53 GYPGA-GGYPAGGYPAGGYPGALSPGGPPAPGGGQFGAPPGAGFGSGYPPQPPAOSYG 111
QY 102 TSPY-----GAPAGPLIVPNLPLPGVVP-----RMLITLITGVKPNANRIALDFQR 149
DB 112 GGPAGQVPVPGFPGGQMPGQY---PGGAPYPSQPSASTGTGQGITLPSN---FDANK 164
QY 150 GNDVAFHFRFNNRNVYCNTKLDNNRGREBROSVPPESGK 193
DB 165 DAELLRKMKFGFTDEQALVDVVSNNRNDROQIKAKFKTWYRK 208

RESULT 39
ID 0804G3 PRELIMINARY; PRT; 485 AA.
AC 0804G3;
DB 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DB Annexin 11b.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Father S.A., Olson E.S., Halpern M.E.;
RT "The Zebrafish Annexin Gene Family.";
RL Genome Res. 0:0-0(2003).
DR EMBL; AY178802; AAC020276.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005544; F:calcium dependent phospholipid binding; IEA.
DR InterPro; IPR001464; Annexin.
DR Pfam; PF00191; annexin; 4.
DR Pfam; PF02162; XYPX; 9.
DR PRINTS; PR00196; ANNEXIN.
DR ProDom; PD000143; Annexin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXIN; 4.
SQ SEQUENCE 485 AA; 51461 MW; 59D4E05C4FA253C CRC64;

Query Match 14.2%; Score 192.5; DB 13; Length 485;
Best Local Similarity 39.7%; Pred. No. 2.4e-08;
Matches 52; Conservative 12; Mismatches 34; Indels 33; Gaps 9;

QY 21 GMPGAMGNQAPAGAGY--PGASVYGYPGQAPP---GAYPGQAPPGAYHAPAGYRGAAPG 76
DB 5 GYPPAGSGYPPASGYPYQPPAGYPPQAGYPPQ---PGAFFPPQGAFF--PQPG 60
QY 77 YVP---GPPSGPAGYPPS-----GQPSA-PGAYATGP-----YGAPAGPL 112
DB 61 AFPAGAGYPPAGGYPAPAGGFPFPQAGYPAAGPAGATPNNPAPAGMGNGHGFAGPAGCM 120
QY 113 IVVP-NLPLPG 122
DB 121 PQGYGVAPAG 131

RESULT 40
ID 071391 PRELIMINARY; PRT; 483 AA.
AC 071391;
DB 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
SQ SEQUENCE FROM N.A.

```

```

DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22386257; PubMed=12477932;
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner I., Scheffer C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schenker C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavani T.L., Scheetz T.E.,
RA Brownstein M.O., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.T.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalski U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=Kidney;
RC Strusberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053208; AAH53208.1; -.
KW Hypothetical protein.
SQ SEQUENCE 483 AA; 51425 MW; E4F7F527E30FEDD7 CRC64;

Query Match 14.1%; Score 191; DB 13; Length 483;
Best Local Similarity 38.4%; Pred. No. 3.3e-08;
Matches 56; Conservative 6; Mismatches 48; Indels 36; Gaps 10;

QY 11 LSGSGNPNQGWPGMAGMNQAGYPPAS--YGYPGQAPP---GAYPGQAPPGAYHAP 66
DB 1 MSYPPGVP-PTGYPPFPFGYPAGGYPYQPPAGYPPQAGYPPQ---PGAFFPPQ 58
QY 67 GAYPGAPAPGYVPSPGSG-----PG-----AYPSGQSPAPG-AYATGPY 105
DB 59 GAFPG--QPGYPSVPSGGMGAPIGLNLPHPGFNASNIQMGPMGMPQNGMGPSPG 115
QY 106 GAPAGPLVVP-NLPLP---GVVPR 127
DB 116 --PPGQMPSTYPNIPATPSGSPSYR 139

RESULT 41
ID 039115 PRELIMINARY; PRT; 177 AA.
AC 039115;
DB 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Glycine and proline rich protein (Genomic DNA, chromosome 5, P1
DE clone:MFC19) (AT5G45350/MFC19_1) (Hypothetical protein).
GN GPRP.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=cv. E.COLI LB392;  
 RX MEDLINE=96189273; PubMed=8605310;  
 RA Matry I., Monfort A., Stiefel V., Ludevid D., Delsen M.,  
 RA Puigdomenech P.;  
 RT "Molecular characterization of the gene coding for GPRP, a class of  
 RT proteins rich in glycine and proline interacting with membranes in  
 RT Arabidopsis thaliana";  
 RL Plant Mol. Biol. 30:625-636(1996).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. E.COLI LB392;  
 RA Puigdomenech P.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=9397451; PubMed=10470850;  
 RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,  
 RA Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.  
 RT Sequence features of the regions of 1,011,550 bp covered by seventeen  
 RT P1 and TAC clones.";  
 RL DNA Res. 6:183-195(1999).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Bath J.,  
 RA Bowser U., Carrinchi P., Dale U.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida U., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai U., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Natusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker U.R.;  
 RT "Arabidopsis cDNA clones.";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RC Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Bath J.,  
 RA Bowser U., Carrinchi P., Chang E., Dale U.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida U., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai U., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Natusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,  
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker U.R.;  
 RT "Arabidopsis ORF clones.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RA Haas B.U., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,  
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 RT annotation.";  
 RL Genome Biol. 0:0-0(2002).  
 RN [17]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RT "Full-length cDNA from Arabidopsis thaliana.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [18]  
 RP EMBL: X68315; CAA59059.1; -;  
 DR EMBL: AB018113; BAB09163.1; -;  
 DR EMBL: AY052250; AAK97720.1; -;  
 DR EMBL: AY060514; AAL31127.1; -;  
 DR EMBL: AY086217; AAM64295.1; -;  
 DR PIR: S65780; S65780.  
 DR InterPro: IPR006031; XYPX.  
 DR Pfam: PF02162; XYPX; 11.  
 KW Hypothetical protein.  
 SQ SEQUENCE 177 AA; 17830 MW; 3EF8094DCAD13F92 CRC64;

Query Match 14.0%; Score 190.5; DB 10; Length 177;  
 Best Local Similarity 44.6%; Pred. No. 1,1e-08;

Matches 50; Conservative 4; Mismatches 41; Indels 17; Gaps 6;  
 QY 13 GSGNPNPGWGMGNQAG--AGYFGASTP--GYGQAPFGAYPGQA--PPGAYHGA 66  
 DB 2 GGDNDNDND--KGFHGYPPAGYPPGAYPPAGYPPQGYP--PPGAYPPAGYPPGAYPPAP 58  
 QY 67 GAYGAPAPGAYPPGPPS-----GPGAYSSGQSPAPGAYATGYPGAPAG 110  
 DB 59 GGYPPAPGAYPPAPGAYPPAGYPPAGHGYPPAGYPPAHSHGAGIGMIMG 110  
 RESULT 42  
 ID 044796 PRELIMINARY; PRT; 329 AA.  
 AC 044796;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN C50D2.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=9906613; PubMed=9651916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Sammons L., Wohldmann P., Bauer C.;  
 RT "The sequence of C. elegans cosmid C50D2.";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF040642; AAB94952.1; -;  
 DR PIR: T32783; T32783.  
 DR WormPep; C50D2.4; CE08889.  
 DR InterPro: IPR008160; Collagen.  
 DR InterPro: IPR002486; Col cuticle\_N.  
 DR Pfam: PF01391; Collagen; 3.  
 DR Pfam: PF01484; Col cuticle\_N; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 329 AA; 32080 MW; 730C7B6101A6BD9A CRC64;  
 Query Match 14.0%; Score 189.5; DB 5; Length 329;  
 Best Local Similarity 43.4%; Pred. No. 2.8e-08;  
 Matches 53; Conservative 1; Mismatches 45; Indels 23; Gaps 7;  
 QY 19 PQCWPGAMGNO--PAGAGGYFGASYPYG----QAPPGAYPGAPGCA-----YHGAPGA 68  
 DB 187 PTCPPPGPKQGGKGNHGPGA--PGTPGNAIPGPPGPGGPGGAUGLPGNPGAPGV 244  
 QY 69 -----YGPAPAGVYPPGPPSGGAYSSGQSPAPGAYATGYPGAPAGPLI--VPYNLPL 120  
 DB 245 PGGVIVPGLTPGA---GPPGPPGPIGPPGQFGAPSSSQPGPGPGGDIIDGAPGNPGS 301  
 QY 121 PG 122  
 DB 302 PG 303  
 RESULT 43  
 Q804G4



```

ID 0804G4 PRELIMINARY; PRT; 526 AA.
AC 0804G4;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Annelin 11a.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Father S.A., Olson E.S., Halpern M.E.;
RT "The Zebrafish Annelin Gene Family.";
RL Genome Res. 0:0-0(2003).
DR EMBL; AY178801; AAC20275.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IEA.
DR InterPro; IPR001464; Annelin.
DR InterPro; IPR006031; XYPX.
DR Pfam; PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXTIN.
DR ProDom; PD000143; Annelin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXTIN; 4.
SQ SEQUENCE 526 AA; 55732 MW; B54010208D8F79F1 CRC64;

Query Match 14.0%; Score 189.5; DB 13; Length 526;
Best Local Similarity 36.8%; Pred. No. 4.9e-08;
Matches 56; Conservative 10; Mismatches 47; Indels 39; Gaps 9;

QY 11 LSGSGNPNQCGMPGAMGNQPAAGAGYPGAS--YRCYPGQAPP--GAYPGQAPPGAYNHAP 66
DB 1 MSYGYPPQSGYPPQGGYPPQPGAXYPPAGYPPQPMYPPQAGYPPQ--PGAYPPQF 58
QY 67 GAYGGAAPGYPPGPPSG-----PGAYPS-----SCQPSAPGAYATGPY--- 105
DB 59 GAFPG--QPGGYPPVPSGSGMGAPIGLDNLPMNPFNASNIGQMANQFADGGAFFNPSWFS 116
QY 106 -----GAPGAPLIV-----PYNL-PLPGSVVR 127
DB 117 GGYGPGQPGGPPAVSPNOPTGMYPQGGGMPQ 148

RESULT 44
Q45904 PRELIMINARY; PRT; 285 AA.
ID 045904
AC 045904;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE W09H1.6b protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z82081; CAB04960.1; -.
DR PIR; T26325; T26325.
DR HSSP; P17931; 1A3K.

```

```

DR WormPep; W09H1.6b; CE16577.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALACTIN; 2.
SQ SEQUENCE 285 AA; 32683 MW; BCB3D635A5848767 CRC64;

Query Match 13.9%; Score 188.5; DB 5; Length 285;
Best Local Similarity 36.4%; Pred. No. 2.8e-08;
Matches 47; Conservative 26; Mismatches 49; Indels 7; Gaps 5;

QY 114 VPYNLPLPGVVRMLITLITGVKNARIALDPQRGN-DVAHF-PPRENNRRVIVCN 171
DB 156 VPYESGLANGLPVGSKLIVGIVEKKARFHNILRKNGDISFHNPRFDEK---VIRN 212
QY 172 TKLDNNWGREERQSVFPESGKPFKIQVLVEPDHFKNV-N-DAHLQVNHRYVKLNEISKL 230
DB 213 SLAANWGNFRRREGNPFKGVGFDLVITQNEEYAFQVFNNGERYISFAHRADP-HDIAGL 271
QY 231 GISGDIDL 239
DB 272 QISGDIELS 280

RESULT 45
Q9NV1 PRELIMINARY; PRT; 277 AA.
ID Q9NV1
AC Q9NV1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Galectin.
OS HCO-GAL-3A.
OC Haemonchus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxId=6289;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20183859; PubMed=10717307;
RA Greenhalgh C.J.; Loukas A.; Donald D.; Nikolaou S.; Newton S.B.;
RT "A family of galectins from haemonchus contortus.";
RL Mol. Biochem. Parasitol. 107:117-121(2000).
DR EMBL; AF105968; AAR63405.1; -.
DR HSSP; P17931; 1A3K.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR001079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALACTIN; 2.
SQ SEQUENCE 277 AA; 31627 MW; 4B4CA7A97982BB82 CRC64;

Query Match 13.8%; Score 187.5; DB 5; Length 277;
Best Local Similarity 37.2%; Pred. No. 3.3e-09;
Matches 48; Conservative 20; Mismatches 54; Indels 7; Gaps 5;

QY 114 VPYNLPLPGVVRMLITLITGVKNARIALDPQRGN-DVAHF-PPRENNRRVIVCN 171
DB 148 VPYESGLANGLPVGSKLIVGIVEKKARFHNILRKNGDISFHNPRFDEK---KAVIRN 204
QY 172 TKLDNNWGREERQSVFPESGKPFKIQVLVEPDHFKNV-N-DAHLQVNHRYVKLNEISKL 230
DB 205 ALAANWGNFRRREGNPFKGVGFDLVITQNEEYAFQVFNNGERYISFAHRADP-QDPNDISGL 263
QY 231 GISGDIDL 239
DB 264 QISGDIEL 272

RESULT 46
Q9NV0 PRELIMINARY; PRT; 283 AA.
ID Q9NV0

```

```

AC Q9NUV0:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Galectin.
GN HCO-GAL-3B.
OS Haemochus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCB1_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2018359; PubMed=10717307;
RA Greenhalgh C.J.; Loukas A.; Donald D.; Nikolaou S.; Newton S.E.;
RT "A family of galectins from haemonchus contortus."
RL Mol. Biochem. Parasitol. 107:117-121(2000).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jiangsu.
RA Li C.H.; Wei X.F.; Li X.R.;
RT "Cloning and expression of galectin gene from Haemonchus contortus."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF105969; AAF63406.1; -
DR EMBL; AY253330; AAP05997.1; -
DR HSSP; P17931; 1A3K.
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
SQ SEQUENCE 283 AA; 32544 MW; F1B2E0DB346BF134 CRC64;

Query Match 13.8%; Score 187.5; DB 5; Length 283;
Best Local Similarity 37.2%; Pred. No. 4.e-08;
Matches 48; Conservative 20; Mismatches 54; Indels 7; Gaps 5;

QY 114 VRYNPLPGGVPRMLITLGVKPNARIALDFQRGN--VAHFH-PRFNNRRVIVCN 171
DB 154 VYESGCIASGFPIDKTLIFGIVEKAKKPNINLRRNDIALHFRPDE--KAVIN 210
QY 172 TLDNNMGREEROSVPEFSGKPEFKIQVLVEPDHFVAVN-DAHLQYNRVKKNLEISKL 230
DB 211 ALAANMGWEEREGKKPEFKVGFDLAIGNAVAFQIFVNGERFTSFAR-QPMDISGL 269
QY 231 GISGIDLT 239
DB 270 QIQGDIET 278

RESULT 47
Q20922 PRELIMINARY; PRT; 313 AA.
AC Q20922;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE F57B1.3 Protein.
GN F57B1.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCB1_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sims M.A.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).

```

```

DR EMBL; Z78064; CAB01509.1; -
DR PIR; J22828; T22828.
DR WormPep; F57B1.3; CP11290.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
SQ SEQUENCE 313 AA; 29902 MW; DSA2C99DF8DCF19E CRC64;

Query Match 13.8%; Score 187.5; DB 5; Length 313;
Best Local Similarity 43.9%; Pred. No. 3.9e-06;
Matches 50; Conservative 2; Mismatches 41; Indels 21; Gaps 6;

QY 14 SGNPNQGPAGMAGNPGAGGYP-----GASYPYGPQAPPYGAAPPYGAAPGA- 68
DB 172 AGPFPSPAPQKG--PGAGVGGQTGASLPPGPAGPSPSPGSGPAGPAGP 229
QY 69 -----YPGAPAPGVYPPSPGYP-----SSGQSPAPAYATGPGAPAGP 111
DB 230 GQVVDVPGTPGAPGPPGP--GPAGAPGQPGSGSGPQPGPG--PDGAGAPGAP 280

RESULT 48
Q96Q57 PRELIMINARY; PRT; 300 AA.
AC Q96Q57;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Beta-galactoside-binding protein galectin-8.
GN LGALS8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Merlier J.; Moisan S.; Potworowski E.F.; St-Pierre Y.;
RT "New galectin-8 isoform specific to NHL."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF037304; AAK69827.1; -
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR01079; Galectin.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR SMART; SM00276; GLECT; 2.
DR PROSITE; PS00309; GALAPTIN; 1.
SQ SEQUENCE 300 AA; 33739 MW; DC26B39C314DE947 CRC64;

Query Match 13.8%; Score 187; DB 4; Length 300;
Best Local Similarity 38.1%; Pred. No. 4.1e-08;
Matches 51; Conservative 16; Mismatches 53; Indels 14; Gaps 7;

QY 111 PLIVPNPL-IGGVPRMLITLGVKPNARIALDFQRGN--VAHFH-PRFNNRR 166
DB 169 PPARLINTMGGGGV-----VKGKFNANASFNVDLAGSKRIALHNPRL--NIK 220
QY 167 VIVCNFKLDNNMGREERQ--SVPEFSGKPEFKIQVLVEPDHFVAVNDAH-LQYNRVK 224
DB 221 AVYRNSFLQESMGEBERNITSPFSGVYFMIIYCDVVEFVAVGVGVSLEVKRFXEL 280
QY 225 NEISKIGISGIDL 238
DB 281 SSIDTLEINGIHL 294

RESULT 49
Q9M0L8 PRELIMINARY; PRT; 179 AA.
AC Q9M0L8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein (Atg19200).

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: August 23, 2004, 14:27:21 ; Search time 59 seconds

(without alignments)  
1178.079 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357  
Sequence: 1 MADNPSLHDAISGSGNPNPQ.....ISKIGISGDIDLTASASYMTI 246

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1357	100.0	246	2 AAW61955	AAW61955 Human gal
2	1357	100.0	246	6 ABU04683	ABU04683 Human exp
3	1315	96.9	250	5 AAU97819	AAU97819 Human cel
4	1315	96.9	250	6 AAG79835	AAG79835 Galectin-
5	1315	96.9	250	6 AAG79836	AAG79836 Human exp
6	1315	96.9	250	6 ABU04685	ABU04685 Human exp
7	1315	96.9	250	6 ABU04677	ABU04677 Human exp
8	1315	96.9	250	6 ABU04676	ABU04676 Human exp
9	1315	96.9	250	6 ADA11009	ADA11009 Human CDN
10	1310	96.5	249	8 ADE77133	ADE77133 Human pro
11	1310	96.5	249	7 ABU04686	ABU04686 Human exp
12	1310	96.5	249	7 ADE56588	ADE56588 Human pro
13	1305	96.2	250	2 AAW71218	AAW71218 Human cel
14	1305	96.2	250	5 AAU97818	AAU97818 Human cel
15	1305	96.2	250	6 ABU04684	ABU04684 Human exp
16	1305	96.2	250	7 ABU04674	ABU04674 Human exp
17	1305	96.2	250	7 ADC53844	ADC53844 Human gal
18	1297	95.6	250	5 AAU97820	AAU97820 Human cel
19	1297	95.6	250	5 AAU97816	AAU97816 Human cel
20	1297	95.6	250	5 AAU97817	AAU97817 Human cel
21	1297	95.6	250	6 ABU04680	ABU04680 Protein d
22	1297	95.6	250	6 ABU04680	ABU04680 Human exp
23	1297	95.6	250	6 ABU04687	ABU04687 Human exp
24	1297	95.6	250	6 ABU04678	ABU04678 Human exp
25	1297	95.6	250	6 ABU04679	ABU04679 Human exp

26	1297	95.6	250	6 ABU04675	ABU04675 Human exp
27	1297	95.6	277	4 AAG75013	AAG75013 Human col
28	1297	95.6	277	6 ABU04681	ABU04681 Human exp
29	1287	94.8	250	6 ABR56317	ABR56317 Human gal
30	1194	88.0	248	2 AAR12532	AAR12532 Human Mac
31	1194	88.0	248	6 ABU04682	ABU04682 Human exp
32	1082	79.7	264	2 AAR12531	AAR12531 Mac2.16 e
33	1082	79.7	264	7 ADC53841	ADC53841 Mouse gal
34	1082	79.7	278	2 AAR13338	AAR13338 Mac-2 pro
35	1079	78.5	262	5 AAU97821	AAU97821 Rat cell
36	1079	78.5	262	7 ADC53847	ADC53847 Rat galec
37	1078	79.4	264	2 AAW71219	AAW71219 Beta-D-ga
38	1078	79.4	264	6 ABU04782	ABU04782 Human exp
39	1075	79.2	262	2 AAR42200	AAR42200 IGE bindi
40	1074	79.1	261	7 ADE56586	ADE56586 Rat Prote
41	1064.5	78.4	263	5 AAU97822	AAU97822 Mouse cel
42	1060	78.1	258	6 AAW61954	AAW61954 Rat galec
43	723	53.3	143	6 AAG79833	AAG79833 N-termina
44	723	53.3	144	6 AAG79834	AAG79834 N-termina
45	710	52.3	141	6 ABR56316	ABR56316 N-termina
46	593	43.7	138	1 AAP60534	AAP60534 C-termina
47	314.5	23.2	341	4 AAB97171	AAB97171 Rainbow t
48	299	22.0	322	7 ADC53842	ADC53842 Mouse gal
49	289	21.3	355	2 AAY06997	AAY06997 Galectin-
50	289	21.3	355	2 AAW85664	AAW85664 Galectin-
51	289	21.3	355	5 ABB77852	ABB77852 Amino aci
52	289	21.3	355	7 ADC53845	ADC53845 Human gal
53	289	21.3	355	7 ADE62929	ADE62929 Human pro
54	289	21.3	355	7 ADD48101	ADD48101 Human pro
55	289	21.3	378	4 AAE13847	AAE13847 Human lun
56	289	21.3	378	7 ADE66747	ADE66747 Human lun
57	289	21.3	378	7 ADE88001	ADE88001 Human lun
58	283	20.9	323	3 AAY56802	AAY56802 Human eos
59	283	20.9	323	5 ABB77853	ABB77853 Amino aci
60	283	20.9	323	5 ABB61494	ABB61494 Human NF-
61	282	20.8	329	5 AAU97036	AAU97036 Human bla
62	277.5	20.4	322	7 ABU63650	ABU63650 Rat urate
63	276.5	20.4	354	7 ADC53848	ADC53848 Rat galec
64	276.5	20.4	354	7 ADE62927	ADE62927 Rat Prote
65	276.5	20.4	354	7 ADD48099	ADD48099 Rat Prote

## ALIGNMENTS

RESULT 1	Query	ID	Description
AAW61955	AAW61955 standard; protein; 246 AA.	XX	
XX	AAW61955;	AC	
XX		XX	
DT	18-SEP-1998 (first entry)		
XX	Human galectin amino acid sequence.		
DE			
XX	Mortalin; galectin; diabetes-mediating protein; insulin; DMF; diabetes; drug screening assay.		
KW			
XX	Homo sapiens.		
OS			
XX	WO9820124-A2.		
XX			
PD	14-MAY-1998.		
XX			
XX	24-OCT-1997; 97WO-IB001627.		
XX			
XX	25-OCT-1996; 96US-0029324P.		
PR	05-NOV-1996; 96US-0030088P.		
PR	05-NOV-1996; 96US-0030186P.		
PR	18-JUL-1997; 97US-00897098.		
XX			
XX	(LARS/) MOSE LARSEN P.		
PA	(FEYS/) FEY S J.		



```
RESULT 3
AAU97819
ID .AAU97819 standard; protein; 250 AA.
XX
AC AAU97819;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human cell membrane anchor protein galectin-3 #4.
XX
KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;
KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;
KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;
KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;
KW polycystic kidney; post-angioplasty restenosis; cytosolic;
KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;
KW vasotropic; hepatotropic; human.
XX
OS Homo sapiens.
XX
PN WO200229031-A2.
XX
PD 11-APR-2002.
XX
PF 01-OCT-2001; 2001WO-IL000918.
XX
PR 04-OCT-2000; 2000US-0237858P.
XX
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
PI Kloog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;
XX
DR WPI; 2002-435333/46.
XX
PT Identifying anchor proteins that bind Ras protein, by producing complexes
PT of Ras and cell membrane proteins in the presence and absence of a Ras
PT antagonist and identifying a complex disrupted by the Ras antagonist.
XX
PS Disclosure; Page 13; 62pp; English.
XX
CC The invention describes a method of identifying cell membrane anchor
CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures
CC where one mixture has a Ras antagonist. A cross linking agent is added,
CC and complexes between Ras protein and other proteins are produced. The
CC complexes are then separated and the proteins binding to Ras are
CC identified. The invention also describes a method useful for identifying
CC drug candidates that inhibit aberrant Ras activity. An antisense compound
CC comprising at least one phosphorothioate-modified nucleotide is useful
CC for disrupting aberrant Ras activity in vivo, by infusing the antisense
CC compound into a patient exhibiting this problem. The method is also
CC useful for identifying anchor proteins for the farnesylated isoforms of H
CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be
CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful
CC for treating diseases characterized by uncontrolled mitosis, including
CC cancers and various non-malignancies such as autoimmune disease (e.g.
CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft
CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty
CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor
CC protein that binds an isoform of Ras
XX
SQ Sequence 250 AA;
Query Match 96.9%; Score 1315; DB 5; Length 250;
Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
```

```
QY 119 PLPGVVPRLMTITIGVKNANRIALDFORGVDVAFHE-PRFENNRVIVCNKTKLDN 177
| | | | |
DB 121 PLPGVVPRLMTITIGVKNANRIALDFORGVDVAFHEPRFENNRVIVCNKTKLDN 180
QY 178 WGREERQSVFPESGKPFKIQVLVEPDHFKVAVNDAAH-LQVNRVYKLINEISKLGISDI 236
| | | | |
DB 181 WGREERQSVFPESGKPFKIQVLVEPDHFKVAVNDAAHLLQVNRVYKLINEISKLGISDI 240
QY 237 DITSASYMTI 246
| | | | |
DB 241 DITSASYMTI 250
QY
DB
RESULT 4
AAG79835
ID AAG79835 standard; protein; 250 AA.
XX
AC AAG79835;
XX
DT 16-APR-2003 (first entry)
XX
DE Galectin-3.
XX
KW Galectin-3; nuclear localisation; BHL domain; Bcl-2; metastasis;
KW N-terminally truncated; polyethylene glycol; tumour; cancer;
KW carbohydrate binding; multimerisation; gene therapy.
XX
OS Homo sapiens.
XX
PN WO2002100343-A2.
XX
PD 19-DEC-2002.
XX
PF 10-JUN-2002; 2002WO-US018478.
XX
PR 08-JUN-2001; 2001US-0296970P.
XX
PA (MAND-) MANDALMED INC.
XX
PI Jarvis GA, John CM, Lefler H;
XX
DR WPI; 2003-183915/18.
XX
PT Novel N-terminally truncated galectin-3 derivatized with polyethylene
PT glycol, useful for treating tumor.
XX
PS Disclosure; Page 23; 88pp; English.
XX
CC The sequences given in AAG79833-35 represent recombinant human galectin-
CC 3's. Galectins bind lactose and the human protein is composed of 250
CC amino acids with a molecular weight of approx. 31000. N-terminally
CC truncated galectin-3 derivatised with at least one molecule of
CC polyethylene glycol is useful for treating a tumour in a patient and for
CC reducing tumour size, and for treating cancer or for preventing
CC metastasis. It is also useful for inhibiting carbohydrate binding and
CC multimerisation of galectin-3, and for inhibiting tumour growth and
CC metastasis in vivo. DNA encoding the truncated galectin-3 is useful in
CC gene therapy
XX
SQ Sequence 250 AA;
Query Match 96.9%; Score 1315; DB 6; Length 250;
Best Local Similarity 98.4%; Pred. No. 1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
```

```
QY 1 MANDFSLHDALSSGNNPQGMWPGAMGNQPAAGAGYPGASYPG-YPCQAPPAGYPGQAPP 59
| | | | |
DB 1 MANDFSLHDALSSGNNPQGMWPGAMGNQPAAGAGYPGASYPGAYPCQAPPAGYPGQAPP 60
QY 60 GAYHGAAGATPGAPAPGVYGPSPSGAGAYSSGQPSAPGAY-ATGPYGAAPGLIVPNYL 118
| | | | |
DB 61 GAYHGAAGATPGAPAPGVYGPSPSGAGAYSSGQPSAPGAYPATGPYGAAPGLIVPNYL 120
```

```

QY 119 PLPGVVRMLITITIGTYKPNANRIALDFORGNDVAEHF-PRFNNRRVIVCNTKLDNN 177
DB 121 PLPGVVRMLITITIGTYKPNANRIALDFORGNDVAEHFNNRRVIVCNTKLDNN 180
QY 178 WREERQSVFPESGKPKIQLVEPDHFKVAVNDAAH-LQYNHRVKLINEISKLIGSDI 236
DB 181 WREERQSVFPESGKPKIQLVEPDHFKVAVNDAAHLLQYNHRVKLINEISKLIGSDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 5
AAG79836
ID AAG79836 standard; protein; 250 AA.
XX
AC AAG79836;
XX
DT 16-APR-2003 (first entry)
XX
DE Galectin-3.
XX
KW Galectin-3; nuclear localisation; BHL domain; Bcl-2; metastasis;
KW N-terminally truncated; polyethylene glycol; tumour; cancer;
KW carbohydrate binding; multimerisation; gene therapy.
XX
OS Homo sapiens.
XX
PN WO2002100343-A2.
XX
PD 19-DEC-2002.
XX
PF 10-JUN-2002; 2002WO-US018478.
XX
PR 08-JUN-2001; 2001US-0296970P.
XX
PI (MAND-) MANDALMED INC.
XX
PI Jarvis GA, John CM, Loeffler H;
XX
DR WPI; 2003-183915/18.
XX
DR N-PSDB; ABA00863.
XX
PT Novel N-terminally truncated galectin-3 derivatized with polyethylene
PT glycol, useful for treating tumor.
XX
PS Disclosure; Page 59; 88pp; English.
XX
XX
CC This sequence shows full length human galectin-3. Galectins bind lactose
CC and the human protein is composed of 250 amino acids with a molecular
CC weight of approx. 31000. N-terminally truncated galectin-3 derivatised
CC with at least one molecule of polyethylene glycol is useful for treating
CC a tumour in a patient and for reducing tumour size, and for treating
CC cancer or for preventing metastasis. It is also useful for inhibiting
CC carbohydrate binding and multimerisation of galectin-3, and for
CC inhibiting tumour growth and metastasis in vivo. DNA encoding the
CC truncated galectin-3 is useful in gene therapy
XX
SQ Sequence 250 AA:

Query Match 96.9%; Score 1315; DB 6; Length 250;
Best Local Similarity 98.4%; Pred. No.1.7e-96;
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
QY 1 MADNFSIHDALSGSGNPNPGQWPGAMGNOPAGAGYFGASYPG-YPGQAPPGAYPGQAPP 59
DB 1 MADNFSIHDALSGSGNPNPGQWPGAMGNOPAGAGYFGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHAGAGAYPGADAPGVYGGPPSGPGAYPSSGQPSAPGAY-ATGPIGAPGAPLIVYNTL 118
DB 61 GAYHAGAGAYPGADAPGVYGGPPSGPGAYPSSGQPSAPGAYPATGPGAPGAPLIVYNTL 120

```

```

QY 119 PLPGVVRMLITITIGTYKPNANRIALDFORGNDVAEHF-PRFNNRRVIVCNTKLDNN 177
DB 121 PLPGVVRMLITITIGTYKPNANRIALDFORGNDVAEHFNNRRVIVCNTKLDNN 180
QY 178 WREERQSVFPESGKPKIQLVEPDHFKVAVNDAAH-LQYNHRVKLINEISKLIGSDI 236
DB 181 WREERQSVFPESGKPKIQLVEPDHFKVAVNDAAHLLQYNHRVKLINEISKLIGSDI 240
QY 237 DLTSASYTMI 246
DB 241 DLTSASYTMI 250

RESULT 6.
ABU04685
ID ABU04685 standard; protein; 250 AA.
XX
AC ABU04685;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human expressed protein tag (EPT) #1351.
XX
KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200278524-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-US009671.
XX
PR 28-MAR-2001; 2001US-0279495P.
PR 21-MAY-2001; 2001US-0292544P.
PR 08-AUG-2001; 2001US-0310801P.
PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0358985P.
XX
PA (ZYCO-) ZYCOS INC.
XX
PI Chicx RM, Tomlinson AJ, Urban RG;
XX
DR WPI; 2003-040607/03.
XX
XX
CC New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
CC cytoskeletal proteins, receptors or transcription factors), useful for
CC treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
CC leukemia.
XX
PT Example 2; SEQ ID NO 1351; 134pp; English.
XX
PS
XX
CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed

```



CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;  
Best Local Similarity 98.4%; Pred. No. 1.7e-96;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSSGSENPDPQMGWGNQPGAGGYPPASYPG-YFGQAPPGAYPGQAP 59  
DB 1 MADNFSLHDALSSGSENPDPQMGWGNQPGAGGYPPASYPGAYPGQAPPGAYPGQAP 60  
QY 60 GAYHGAAPGAYPGAPGAYPPGPGSGGAYPPSGQSPAPGAY-ATGPPYGAAPGLIYVYNL 118  
DB 61 GAYHGAAPGAYPGAPGAYPPGPGSGGAYPPSGQSPAPGAYPATGPPYGAAPGLIYVYNL 120  
QY 119 PLPGGVVPRMLITILGTVPKNANRIALDFQGNDAVAFHFNPRNENRRVIVCNTKLDNN 177  
DB 121 PLPGGVVPRMLITILGTVPKNANRIALDFQGNDAVAFHFNPRNENRRVIVCNTKLDNN 180  
QY 178 WGREERQSVPPESGKPFKIOVLVEPDHFKVAVNDAA-LQYNHRVKKLNEISKLGISDI 236  
DB 181 WGREERQSVPPESGKPFKIOVLVEPDHFKVAVNDAAHLQYNHRVKKLNEISKLGISDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 7  
ABU04677  
ID ABU04677 standard; protein: 250 AA.  
XX  
AC ABU04677;  
XX  
DT 29-JAN-2003 (first entry)  
XX  
DE Human expressed protein tag (EPT) #1343.  
XX  
KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
KW protease; protease inhibitor; transporter; cytoskeletal protein;  
KW receptor; transcription factor; cancer; MHC;  
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200278524-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 28-MAR-2002; 2002WO-US009671.  
XX  
PR 28-MAR-2001; 2001US-0279495P.  
PR 21-MAY-2001; 2001US-0292544P.  
PR 08-AUG-2001; 2001US-0310801P.  
PR 01-OCT-2001; 2001US-0326370P.  
PR 04-DEC-2001; 2001US-0336780P.  
PR 20-FEB-2002; 2002US-0358985P.  
XX  
PA (ZYCO-) ZYCOS INC.  
XX  
PI Chiciz RM, Tomlinson AJ, Urban RG;  
XX  
DR WPI; 2003-040607/03.  
XX  
PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
PT leukaemia.  
XX  
PS Example 2; SEQ ID NO 1343; 134P; English.

XX  
CC The invention describes a purified polypeptide, which comprises a  
CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
CC transporter, cytoskeletal protein, receptor or transcription factor. The  
CC polypeptide is useful as an immunogenic composition for eliciting in a  
CC mammal an immunogenic response directed against any of the purified  
CC polypeptide. The purified polypeptide, or the antibody that binds to this  
CC polypeptide, is useful for treating cancer. The polypeptide is also  
CC useful for identifying compounds that binds to a naturally processed  
CC class I or class II MHC-binding polypeptide. The polypeptides and  
CC polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukaemia. These are also useful for screening agents for  
CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling. Note: This sequence does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;  
Best Local Similarity 98.4%; Pred. No. 1.7e-96;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSSGSENPDPQMGWGNQPGAGGYPPASYPG-YFGQAPPGAYPGQAP 59  
DB 1 MADNFSLHDALSSGSENPDPQMGWGNQPGAGGYPPASYPGAYPGQAPPGAYPGQAP 60  
QY 60 GAYHGAAPGAYPGAPGAYPPGPGSGGAYPPSGQSPAPGAY-ATGPPYGAAPGLIYVYNL 118  
DB 61 GAYHGAAPGAYPGAPGAYPPGPGSGGAYPPSGQSPAPGAYPATGPPYGAAPGLIYVYNL 120  
QY 119 PLPGGVVPRMLITILGTVPKNANRIALDFQGNDAVAFHFNPRNENRRVIVCNTKLDNN 177  
DB 121 PLPGGVVPRMLITILGTVPKNANRIALDFQGNDAVAFHFNPRNENRRVIVCNTKLDNN 180  
QY 178 WGREERQSVPPESGKPFKIOVLVEPDHFKVAVNDAA-LQYNHRVKKLNEISKLGISDI 236  
DB 181 WGREERQSVPPESGKPFKIOVLVEPDHFKVAVNDAAHLQYNHRVKKLNEISKLGISDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 8  
ABU04676  
ID ABU04676 standard; protein: 250 AA.  
XX  
AC ABU04676;  
XX  
DT 29-JAN-2003 (first entry)  
XX  
DE Human expressed protein tag (EPT) #1342.  
XX  
KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
KW protease; protease inhibitor; transporter; cytoskeletal protein;  
KW receptor; transcription factor; cancer; MHC;  
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200278524-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 28-MAR-2002; 2002WO-US009671.  
XX  
PR 28-MAR-2001; 2001US-0279495P.  
PR 21-MAY-2001; 2001US-0292544P.  
PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 XX (ZYCO-) ZYCO INC.  
 XX Chicz RM, Tomlinson AJ, Urban RG;  
 XX WPI; 2003-040607/03.  
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 PS Example 2; SEQ ID NO 1342; 134p; English.  
 XX The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (BPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;

Best Local Similarity 98.4%; Pred. No. 1.7e-96;

Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MNDNPSLHDLSSGSGNPNQMGWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 1 MANDFSLHDLSSGSGNPNQMGWGNOPAGAGYPGASYPGAYPGQAPP 60  
 QY 60 GAYHGAPGAYPGAPAGYVPGPSGPGAYSSGQPSAPGAY-ATGPGAPAGLIVPYNL 118  
 DB 61 GAYHGAPGAYPGAPAGYVPGPSGPGAYSSGQPSAPGAYPATGPGAPAGLIVPYNL 120  
 QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHF-PRFENNRRVIVCNTKLDNN 177  
 DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHFPRFENNRRVIVCNTKLDNN 180  
 QY 178 WGEREESQVFPFSSGKPFKIQLVVEPDHFKVAVNDAH-LQYNHRVKKLNISKLGISGDI 236  
 DB 181 WGEREESQVFPFSSGKPFKIQLVVEPDHFKVAVNDAHALQYNHRVKKLNISKLGISGDI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250

RESULT 9

ADAL1009 standard; protein; 250 AA.

AC ADAL1009;

XX 06-NOV-2003 (first entry)

XX Human cDNA differentially expressed in colon cancer #85 product.

XX differential expression; colon cancer; cancer; human.

XX Homo sapiens.  
 OS  
 XX US2002160382-A1.  
 PN  
 XX 31-OCT-2002.  
 PD  
 XX 11-OCT-2001; 2001US-00981353.  
 XX  
 XX 11-OCT-2000; 2000US-0239841P.  
 PR  
 XX (LASEK/) LASEK A W.  
 PA (LASEK/) JONES D A.  
 XX  
 XX Lasek AW, Jones DA;  
 PI  
 XX WPI; 2003-265756/26.  
 DR N-PSDB; ADAL1008.  
 DR  
 XX New combination comprising cDNAs that are differentially expressed in  
 PT colon disorder, useful for diagnosing, treating, staging or monitoring  
 PT treatment for colon cancers.  
 PT  
 PS Example 14; SEQ ID NO 127; 231p; English.

XX The invention relates to a combination comprising cDNAs that are  
 CC differentially expressed in colon disorder. The methods and compositions  
 CC of the present invention are useful for diagnosing, treating, staging or  
 CC monitoring treatment for colon cancer. They are also useful in high  
 CC throughput methods for using cDNAs to detect differential expression of  
 CC nucleic acids in a sample, screening molecules or compounds to identify a  
 CC ligand which specifically binds a cDNA and using a protein to screen  
 CC molecules or compounds to identify at least one ligand which specifically  
 CC binds the protein. The present sequence represents the amino acid  
 CC sequence of a human cDNA differentially expressed in colon cancer  
 CC protein.

XX Sequence 250 AA;

Query Match 96.9%; Score 1315; DB 6; Length 250;

Best Local Similarity 98.4%; Pred. No. 1.7e-96;

Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MNDNPSLHDLSSGSGNPNQMGWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 1 MANDFSLHDLSSGSGNPNQMGWGNOPAGAGYPGASYPGAYPGQAPP 60  
 QY 60 GAYHGAPGAYPGAPAGYVPGPSGPGAYSSGQPSAPGAY-ATGPGAPAGLIVPYNL 118  
 DB 61 GAYHGAPGAYPGAPAGYVPGPSGPGAYSSGQPSAPGAYPATGPGAPAGLIVPYNL 120  
 QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHF-PRFENNRRVIVCNTKLDNN 177  
 DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHFPRFENNRRVIVCNTKLDNN 180  
 QY 178 WGEREESQVFPFSSGKPFKIQLVVEPDHFKVAVNDAH-LQYNHRVKKLNISKLGISGDI 236  
 DB 181 WGEREESQVFPFSSGKPFKIQLVVEPDHFKVAVNDAHALQYNHRVKKLNISKLGISGDI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250

RESULT 10

ADE77133 standard; protein; 250 AA.

XX ADE77133;

XX 29-JAN-2004 (first entry)

XX Human protein expressed in a liver disorder #77.

XX human; liver disorder; hyperlipidaemia; hypertension; type II diabetes;  
KM tumour; liver; inflammatory disorder; immune response disorder;  
KM high-throughput screening; differential gene expression; gene therapy.  
OS Homo sapiens.  
XX US2003108871-A1.  
PN 12-JUN-2003.  
PD 30-JUL-2001; 2001US-00919039.  
XX 28-JUL-2000; 2000US-0222113P.  
PR (KASE/) KASER M R.  
XX Kaser MR;  
PI WPI: 2004-031227/03.  
DR N-PSDB; ADE77132.  
XX Composition comprising several cDNAs that are differentially expressed in  
PT treated human C3A liver cell cultures, useful for treating liver  
PT disorders.  
XX Claim 1; SEQ ID NO 298, 41pp; English.  
PS The invention relates to a composition comprising several cDNAs that are  
CC differentially expressed in a liver disorder. The composition is useful  
CC for treating liver disorder such as hyperlipidaemia, hypertension, type  
CC II diabetes, tumours of the liver and disorders of the inflammatory and  
CC immune response. The composition is useful for a high-throughput method  
CC of screening several molecules or compounds to identify a ligand which  
CC specifically binds a cDNA. A protein encoded by the cDNA is useful for a  
CC high-throughput method for using a protein to screen several molecules or  
CC compounds to identify at least one ligand which specifically binds the  
CC protein which involves combining the protein encoded by the cDNA with  
CC several of molecules or compounds under conditions to allow specific  
CC binding, and detecting specific binding between the protein and a  
CC molecule or compound, therefore identifying a ligand which specifically  
CC binds the protein. The composition is useful for detecting and  
CC quantifying differential gene expression, can be used in gene therapy, to  
CC formulate prognosis and to design a treatment regimen and to monitor the  
CC efficacy of treatment. The present sequence represents the amino acid  
CC sequence of a protein encoded by a cDNA differentially expressed in a  
CC liver disorder.  
SQ Sequence 250 AA;  
Query Match 96.9%; Score 1315; DB 8; Length 250;  
Best Local Similarity 98.4%; Pred. No. 1.7e-96;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPG-YEQAPPAGYPPGAPP 59  
DB 1 MADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPG-YEQAPPAGYPPGAPP 60  
QY 60 GAYHGAAPGAPGAPAPGVPPGPPSGPAGYSSGQPSAPGAY-ATGPGAGAGPLIVPYNL 118  
DB 61 GAYHGAAPGAPGAPAPGVPPGPPSGPAGYSSGQPSAPGAYPATGPGAGAGPLIVPYNL 120  
QY 119 PLTGGVPPRLITLITLGTAKENANRIALDPORGNDVAFHF-PRFENNRRVIVCQTKLDNN 177  
DB 121 PLTGGVPPRLITLITLGTAKENANRIALDPORGNDVAFHFPRFENNRRVIVCQTKLDNN 180  
QY 178 WGREROSVFPFSSGKPEKIQVLVEPDHFKVAVNDAA-LOYNRRVKLANEISKLGISGDI 236  
DB 181 WGREROSVFPFSSGKPEKIQVLVEPDHFKVAVNDAALOYNRRVKLANEISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 11  
ID AB004686 standard; protein; 249 AA.  
XX AB004686;  
AC 29-JAN-2003 (first entry)  
XX 29-JAN-2003 (first entry)  
DE Human expressed protein tag (EPT) #1352.  
XX Translational profiling: expressed protein tag; EPT; kinase; phosphatase;  
KM protease; protease inhibitor; transporter; cytoskeletal protein;  
KM receptor; transcription factor; cancer; MHC;  
KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
XX Homo sapiens.  
OS WO200278524-A2.  
XX WO200278524-A2.  
PN 10-OCT-2002.  
PD 28-MAR-2002; 2002WO-US009671.  
XX 28-MAR-2001; 2001US-0279495P.  
PR 21-MAY-2001; 2001US-0292544P.  
PR 08-AUG-2001; 2001US-0310801P.  
PR 01-OCT-2001; 2001US-0316370P.  
PR 04-DEC-2001; 2001US-0336780P.  
PR 20-FEB-2002; 2002US-0358985P.  
XX (ZYCO-) ZYCO INC.  
PI Chicx RM, Tomlinson AJ, Urban RG;  
DR WPI: 2003-040607/03.  
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
PT leukemia.  
PS Example 2; SEQ ID NO 1352; 134pp; English.  
XX The invention describes a purified polypeptide, which comprises a  
CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
CC transporter, cytoskeletal protein, receptor or transcription factor. The  
CC polypeptide is useful as an immunogenic composition for eliciting in a  
CC mammal an immunogenic response directed against any of the purified  
CC polypeptide. The purified polypeptide, or the antibody that binds to this  
CC polypeptide, is useful for treating cancer. The polypeptide is also  
CC useful for identifying compounds that binds to a naturally processed  
CC class I or class II MHC-binding polypeptide. The polypeptides and  
CC polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukaemia. These are also useful for screening agents for  
CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling. Note: This sequence does not appear in the printed  
CC ftp.wipo.int/pub/published\_pcl\_sequences  
SQ Sequence 249 AA;  
Query Match 96.5%; Score 1310; DB 6; Length 249;  
Best Local Similarity 98.4%; Pred. No. 4.3e-96;  
Matches 245; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 2 ADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPG-YEQAPPAGYPPGAPP 60  
DB 1 ADNFSIHDALSGSGNENPQGWPGAMGNOPAGAGGYGASYPG-YEQAPPAGYPPGAPP 60

QY 61 AYHGAPGAYPGAPAGVYPPSPSGGAYPSSGQPSAPGAY-ATGPGYAPAGPLIYVYMLP 119  
 DB 61 AYHGAPGAYPGAPAGVYPPSPSGGAYPSSGQPSAPGAYPATGPGYAPAGPLIYVYMLP 120  
 QY 120 LPGAIVPMLITLIGTVKPNANRILADPQGNDAVFHF-PPENNNRRVIVCNTKLDNNW 178  
 DB 121 LPGAIVPMLITLIGTVKPNANRILADPQGNDAVFHF-PPENNNRRVIVCNTKLDNNW 180  
 QY 179 GREERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGID 237  
 DB 181 GREERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGID 240  
 QY 238 LTSASYTMI 246  
 DB 241 LTSASYTMI 249

RESULT 12  
 ADE56588 standard; protein; 249 AA.  
 AC ADE56588;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human Protein P17931, SEQ ID NO 2442.  
 XX  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SN1; Chung.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003016475-A2.  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-033347P.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 PI Woolf C, D'ureo D, Befort K, Costrigan M;  
 DR WPI; 2003-268312/26.  
 DR GENBANK; P17931.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SN1)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_ptc\_sequences.  
 XX  
 SQ Sequence 249 AA;  
 XX

Query Match 96.5%; Score 1310; DB 7; Length 249;  
 Best Local Similarity 98.4%; Pred. No. 4.3e-96;  
 Matches 245; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 2 ADNFSIHDALSGSGNPNQGMFGAWGNQPAAGGYPGASVYRG-YPGQAPPGAYPQQAAPG 60  
 DB 1 ADNFSIHDALSGSGNPNQGMFGAWGNQPAAGGYPGASVYPGQAPPGAYPQQAAPG 60  
 QY 61 AYHGAPGAYPGAPAGVYPPSPSGGAYPSSGQPSAPGAY-ATGPGYAPAGPLIYVYMLP 119  
 DB 61 AYHGAPGAYPGAPAGVYPPSPSGGAYPSSGQPSAPGAYPATGPGYAPAGPLIYVYMLP 120  
 QY 120 LPGAIVPMLITLIGTVKPNANRILADPQGNDAVFHF-PPENNNRRVIVCNTKLDNNW 178  
 DB 121 LPGAIVPMLITLIGTVKPNANRILADPQGNDAVFHF-PPENNNRRVIVCNTKLDNNW 180  
 QY 179 GREERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGID 237  
 DB 181 GREERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAH-LQYHRYVKLNEISKLGISGID 240  
 QY 238 LTSASYTMI 246  
 DB 241 LTSASYTMI 249

RESULT 13  
 AAW71218 standard; protein; 250 AA.  
 ID AAW71218  
 XX  
 AC AAW71218;  
 XX  
 DT 30-OCT-1998 (first entry)  
 XX  
 DE Beta-D-galactoside-binding protein designated L-31-gal-lectin.  
 XX  
 KW Beta-D-galactoside-binding protein; L-31-gal-lectin;  
 KW metastatic potential; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US5801002-A.  
 PD 01-SEP-1998.  
 XX  
 PF 22-NOV-1995; 95US-00562311.  
 XX  
 PR 06-JAN-1989; 89US-00294249.  
 PR 05-APR-1991; 91US-00681242.  
 PR 26-JAN-1994; 94US-00188225.  
 XX  
 PA (KARM-) KARMANOS CANCER INST BARBARA ANN.  
 PI Raz A;  
 XX  
 DR WPI; 1998-494766/42.  
 DR N-PSDB; AAV54735.  
 XX  
 PT Test for metastatic potential of cell sample - by measuring binding of  
 PT antibody to L-31-gal-lectin on cell surface.

XX Claim 1; Fig 6A-B; 24pp; English.  
 PS  
 XX  
 CC The present sequence represents a beta-D-galactoside-binding protein  
 CC designated L-31-gal-lectin. The sequence is derived from clone 1. The  
 CC specification describes a method for testing a cell sample for metastatic  
 CC potential. The method comprises contacting the sample with a labelled  
 CC antibody that binds to endogenous cell-surface L-31-gal-lectin, removing  
 CC unbound antibody, and determining the amount of bound antibody as a  
 CC measure of L-31-gal-lectin expression, where the metastatic potential  
 CC increases as the level of L-31-gal-lectin expression increases. The  
 CC antibody is produced by immunisation with a L-31-gal-lectin protein  
 XX  
 SQ Sequence 250 AA;  
 Query Match 96.2%; Score 1305; DB 2; Length 250;  
 Best Local Similarity 97.6%; Pred. No. 1.1e-95;  
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;  
 QY 1 MADNFSLHDALSSGSPNPQGWPGAMGNQPAAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 1 MADNFSLHDALSSGSPNPQGWPGAMGNQPAAGAGYPGASYPGAYPGQAPPAYPGQAPP 60  
 QY 60 GAYHGAAPGAYPGAPAPGAYPGSPSGGAYPGSSGQSPASAGAY-ATGPGYAPAPGLIYPPYNTL 118  
 DB 61 GAYHGAAPGAYPGAPAPGAYPGSPSGGAYPGSSGQSPASAGAYATGPGYAPAPGLIYPPYNTL 120  
 QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHF-PRFNNRNRRVIVCNTKLDNN 177  
 DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHFNNRRVIVCNTKLDNN 180  
 QY 178 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDAN-LQYNHRVKKINEISKLGISGDI 236  
 DB 181 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDANHLQYNHRVKKINEIRKLGISGDI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250  
 RESULT 14  
 ID AAU97818 standard; protein; 250 AA.  
 AC AAU97818;  
 DT 27-AUG-2002 (first entry)  
 DE Human cell membrane anchor protein galectin-3 #3.  
 KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;  
 KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;  
 KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;  
 KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;  
 KW polycystic kidney; post-angioplasty restenosis; cytoskeletal;  
 KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;  
 KW vasotropic; hepatotropic; human.  
 OS Homo sapiens.  
 PN WO200229031-A2.  
 PD 11-APR-2002.  
 PP 01-OCT-2001; 2001WO-IL000918.  
 PR 04-OCT-2000; 2000US-0237858P.  
 RA (UNRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 PI KIoog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;  
 DR WPI; 2002-435333/46.

XX  
 PT Identifying anchor proteins that bind Ras protein, by producing complexes  
 PT of Ras and cell membrane proteins in the presence and absence of a Ras  
 PT antagonist and identifying a complex disrupted by the Ras antagonist.  
 XX  
 PS Disclosure; Page 13; 62pp; English.  
 XX  
 CC The invention describes a method of identifying cell membrane anchor  
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures  
 CC where one mixture has a Ras antagonist. A cross linking agent is added,  
 CC and complexes between Ras protein and other proteins are produced. The  
 CC complexes are then separated and the proteins binding to Ras are  
 CC identified. The invention also describes a method useful for identifying  
 CC drug candidates that inhibit aberrant Ras activity. An antisense compound  
 CC comprising at least one phosphorothioate-modified nucleotide is useful  
 CC for disrupting aberrant Ras activity in vivo, by infusing the antisense  
 CC compound into a patient exhibiting this problem. The method is also  
 CC useful for identifying anchor proteins for the farnesylated isoforms of H  
 CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be  
 CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful  
 CC for treating diseases characterised by uncontrolled mitosis, including  
 CC cancers and various non-malignancies such as autoimmune disease (e.g.  
 CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft  
 CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty  
 CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor  
 CC protein that binds an isoform of Ras  
 XX  
 SQ Sequence 250 AA;  
 Query Match 96.2%; Score 1305; DB 5; Length 250;  
 Best Local Similarity 98.0%; Pred. No. 1.1e-95;  
 Matches 245; Conservative 0; Mismatches 1; Indels 4; Gaps 4;  
 QY 1 MADNFSLHDALSSGSPNPQGWPGAMGNQPAAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 1 MADNFSLHDALSSGSPNPQGWPGAMGNQPAAGAGYPGASYPGAYPGQAPPAYPGQAPP 60  
 QY 60 GAYHGAAPGAYPGAPAPGAYPGSPSGGAYPGSSGQSPASAGAY-ATGPGYAPAPGLIYPPYNTL 118  
 DB 61 GAYHGAAPGAYPGAPAPGAYPGSPSGGAYPGSSGQSPASAGAYATGPGYAPAPGLIYPPYNTL 120  
 QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHF-PRFNNRNRRVIVCNTKLDNN 177  
 DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVFHFNNRRVIVCNTKLDNN 180  
 QY 178 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDAN-LQYNHRVKKINEISKLGISGDI 236  
 DB 181 WGERERQSVFPFESGKPKIQLVVEPDHFKVAVNDANHLQYNHRVKKINEIRKLGISGDI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250  
 RESULT 15  
 ID ABU04684 standard; protein; 250 AA.  
 AC ABU04684;  
 DT 29-JAN-2003 (first entry)  
 DE Human expressed protein tag (EPT) #1350.  
 KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
 OS Homo sapiens.  
 PN WO200278524-A2.



QY 119 PLPGVVPRLITLITIGVKNANRIALDPORGNDVAFHF-PRENNRRVIVCNTKLDNN 177  
 DB 121 PLPGVVPRLITLITIGVKNANRIALDPORGNDVAFHFENRRVIVCNTKLDNN 180  
 QY 178 WGEERQSVFPFESGKPFKIQVLEPDHFVAVNDAH-LQYNNRVKKNLEISLGISSGI 236  
 DB 181 WGEERQSVFPFESGKPFKIQVLEPDHFVAVNDAHLLQYNNRVKKNLEISLGISSGI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250

RESULT 17  
 ADCS3844  
 ID ADCS3844 standard; protein; 250 AA.  
 XX  
 AC ADCS3844;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Human galectin 3 protein.  
 XX  
 KW galectin-3; nephritis; glomerular nephritis; antiinflammatory;  
 KW glomerular infiltration; apoptosis; human.  
 XX  
 OS Homo sapiens.  
 XX  
 EN JP2002322082-A.  
 PD 08-NOV-2002.  
 XX  
 PF 26-APR-2001; 2001JP-00129200.  
 XX  
 PR 26-APR-2001; 2001JP-00129200.  
 XX  
 PA (PROT-) PROTEGENE KK.  
 XX  
 DR WPI; 2003-367092/35.  
 XX  
 PT Agents for prevention and treatment of nephritis, comprise galectin-1,  
 PT galectin-3, or galectin-9, by inhibition of intraglomerular infiltration  
 PT of leukocytes, CD8 positive cells, and induction of apoptosis of CD8  
 PT positive cells.  
 XX  
 PS Disclosure; SEQ ID NO 5; 31pp; Japanese.  
 XX  
 CC This invention relates to the use of novel mammal derived galectin-1  
 CC (G1), -3 (G3) and -9 (G9) proteins as effective ingredients for  
 CC prevention and treatment of nephritis. The invention discloses agents for  
 CC prevention and treatment of nephritis, particularly glomerular nephritis  
 CC and may have antiinflammatory activities. The method of the invention  
 CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and  
 CC apoptosis of CD8 positive cells. The method and sequences of the  
 CC invention may be used for prevention and treatment of nephritis,  
 CC particularly glomerular nephritis including inhibition of glomerular  
 CC infiltration of leukocytes, CD8 positive cells and apoptosis of CD8  
 CC positive cells. The present sequence represents the human galectin 3  
 CC protein of the invention.  
 XX  
 SQ Sequence 250 AA;

Query Match 96.2%; Score 1305; DB 7; Length 250;  
 Best Local Similarity 98.0%; Pred. No. 1,1e-95;  
 Matches 245; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

QY 1 MADNFSJHDLALSSGNGNPGWPGAWGNOPAGAGYVPGASVPG-YPQAGAPPGAYPGQAPP 59  
 DB 1 MADNFSJHDLALSSGNGNPGWPGAWGNOPAGAGYVPGASVPGAYPGQAPP 60  
 QY 60 GAYHAGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 118  
 DB 61 GAYHAGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 120

QY 119 PLPGVVPRLITLITIGVKNANRIALDPORGNDVAFHF-PRENNRRVIVCNTKLDNN 177  
 DB 121 PLPGVVPRLITLITIGVKNANRIALDPORGNDVAFHFENRRVIVCNTKLDNN 180  
 QY 178 WGEERQSVFPFESGKPFKIQVLEPDHFVAVNDAH-LQYNNRVKKNLEISLGISSGI 236  
 DB 181 WGEERQSVFPFESGKPFKIQVLEPDHFVAVNDAHLLQYNNRVKKNLEISLGISSGI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250

RESULT 18  
 AAU97820  
 ID AAU97820 standard; protein; 250 AA.  
 XX  
 AC AAU97820;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE Human cell membrane anchor protein galectin-3 #5.  
 XX  
 KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;  
 KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;  
 KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;  
 KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;  
 KW polycystic kidney; post-angioplasty restenosis; cytostatic;  
 KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;  
 KW vasotropic; hepatotropic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 EN WO200229031-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 01-OCT-2001; 2001WO-IL000918.  
 XX  
 PR 04-OCT-2000; 2000US-0237858P.  
 XX  
 RR (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX  
 PA Klooog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;  
 PI  
 PL Klooog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;  
 DR WPI; 2002-435333/46.  
 XX  
 PT Identifying anchor proteins that bind Ras protein, by producing complexes  
 PT of Ras and cell membrane proteins in the presence and absence of a Ras  
 PT antagonist and identifying a complex disrupted by the Ras antagonist.  
 XX  
 PS Disclosure; Page 14; 62pp; English.  
 XX  
 CC The invention describes a method of identifying cell membrane anchor  
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures  
 CC where one mixture has a Ras antagonist. A cross linking agent is added,  
 CC and complexes between Ras protein and other proteins are produced. The  
 CC complexes are then separated and the proteins binding to Ras are  
 CC identified. The invention also describes a method useful for identifying  
 CC drug candidates that inhibit aberrant Ras activity. An antisense compound  
 CC comprising at least one phosphorothioate-modified nucleotide is useful  
 CC for disrupting aberrant Ras activity in vivo, by infusing the antisense  
 CC compound into a patient exhibiting this problem. The method is also  
 CC useful for identifying anchor proteins for the farnesylated isoforms of H  
 CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be  
 CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful  
 CC for treating diseases characterized by uncontrolled mitosis, including  
 CC cancers and various non-malignancies such as autoimmune disease (e.g.  
 CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft  
 CC rejection, atherosclerosis, polycystic kidneys and post-angioplasty  
 CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor  
 CC protein that binds an isoform of Ras

XX S0 Sequence 250 AA;  
 Query Match 95.6%; Score 1297; DB 5; Length 250;  
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;  
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNLSLHDALSGSGNPQGMWGNQAPAGAGYPCASYPG-YPGQAPPGAYGQAP 59  
 DB 1 MADNLSLHDALSGSGNPQGMWGNQAPAGAGYPCASYPGAYPGQAPPGAYGQAP 60  
 QY 60 GAYHAPGAYPGAPAPGAYPPGSPGAYPSSGQSPAPGAY-ATGYPGAPAGLIVPNL 118  
 DB 61 GAYPAPGAYPGAPAPGAYPPGSPGAYPSSGQSPATGAYPATGYPGAPAGLIVPNL 120  
 QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFNH-PRFNENNRVIVCNTKLDNN 177  
 DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFNHPRFNENNRVIVCNTKLDNN 180  
 QY 178 WGERERQSYFPESGKPEFKIQVLVEPDHFKVAVNDAN-LQYNHRYVKLNEISKLGSDI 236  
 DB 181 WGERERQSYFPESGKPEFKIQVLVEPDHFKVAVNDANHLQYNHRYVKLNEISKLGSDI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250

RESULT 19  
 AAU97816  
 ID AAU97816 standard; protein; 250 AA.  
 XX AAU97816;  
 DT 27-AUG-2002 (first entry)  
 XX Human cell membrane anchor protein galectin-3 #1.  
 DE Galectin-3; cell membrane anchor protein; Ras; antisense technology;  
 XX farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;  
 KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;  
 KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;  
 KW polycystic kidney; post-angioplasty restenosis; cytostatic;  
 KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;  
 KW vasotropic; hepatotropic; human.  
 OS Homo sapiens.  
 XX  
 PN WC0200229031-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 01-OCT-2001; 2001WO-11000918.  
 XX  
 PR 04-OCT-2000; 2000US-0237858P.  
 XX  
 PA (UTRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 PI Kluog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;  
 XX  
 DR WPI: 2002-435333/46.  
 XX  
 DR N-PSDB; ABK52348.  
 XX  
 XX Identifying anchor proteins that bind Ras protein, by producing complexes  
 PT of Ras and cell membrane proteins in the presence and absence of a Ras  
 PT antagonist and identifying a complex disrupted by the Ras antagonist.  
 XX  
 PS Disclosure; Page 11-12; 62pp; English.  
 CC The invention describes a method of identifying cell membrane anchor  
 CC proteins that bind a Ras protein, involving preparing 2 reaction mixtures  
 CC where one mixture has a Ras antagonist. A cross linking agent is added,  
 CC and complexes between Ras protein and other proteins are produced. The

CC complexes are then separated and the proteins binding to Ras are  
 CC identified. The invention also describes a method useful for identifying  
 CC drug candidates that inhibit aberrant Ras activity. An antisense compound  
 CC comprising at least one phosphorothioate-modified nucleotide is useful  
 CC for disrupting aberrant Ras activity in vivo, by infusing the antisense  
 CC compound into a patient exhibiting this problem. The method is also  
 CC useful for identifying anchor proteins for the farnesylated isoforms of H  
 CC -Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be  
 CC oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful  
 CC for treating diseases characterised by uncontrolled mitosis, including  
 CC cancers and various non-malignancies such as autoimmune disease (e.g.  
 CC type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft  
 CC rejection, atherosclerosis, polycystic kidney and post-angioplasty  
 CC restenosis. This sequence encodes a galectin-3, a cell-membrane anchor  
 CC protein that binds an isoform of Ras

XX S0 Sequence 250 AA;  
 Query Match 95.6%; Score 1297; DB 5; Length 250;  
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;  
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNLSLHDALSGSGNPQGMWGNQAPAGAGYPCASYPG-YPGQAPPGAYGQAP 59  
 DB 1 MADNLSLHDALSGSGNPQGMWGNQAPAGAGYPCASYPGAYPGQAPPGAYGQAP 60  
 QY 60 GAYHAPGAYPGAPAPGAYPPGSPGAYPSSGQSPAPGAY-ATGYPGAPAGLIVPNL 118  
 DB 61 GAYPAPGAYPGAPAPGAYPPGSPGAYPSSGQSPATGAYPATGYPGAPAGLIVPNL 120  
 QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFNH-PRFNENNRVIVCNTKLDNN 177  
 DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFNHPRFNENNRVIVCNTKLDNN 180  
 QY 178 WGERERQSYFPESGKPEFKIQVLVEPDHFKVAVNDAN-LQYNHRYVKLNEISKLGSDI 236  
 DB 181 WGERERQSYFPESGKPEFKIQVLVEPDHFKVAVNDANHLQYNHRYVKLNEISKLGSDI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250

RESULT 20  
 AAU97817  
 ID AAU97817 standard; protein; 250 AA.  
 XX AAU97817;  
 DT 27-AUG-2002 (first entry)  
 XX Human cell membrane anchor protein galectin-3 #2.  
 DE Galectin-3; cell membrane anchor protein; Ras; antisense technology;  
 KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;  
 KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;  
 KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;  
 KW polycystic kidney; post-angioplasty restenosis; cytostatic;  
 KW immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;  
 KW vasotropic; hepatotropic; human.  
 OS Homo sapiens.  
 XX  
 PN WC0200229031-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 01-OCT-2001; 2001WO-11000918.  
 XX  
 PR 04-OCT-2000; 2000US-0237858P.  
 XX  
 PA (UTRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX





KM protease; protease inhibitor; transporter; cytoskeletal protein;  
 KM receptor; transcription factor; cancer; MHC;  
 KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
 XX Homo sapiens.  
 OS  
 XX WO200278524-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 28-MAR-2002; 2002MO-US009671.  
 XX  
 XX 28-MAR-2001; 2001US-0279495P.  
 XX 21-MAY-2001; 2001US-0292544P.  
 PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 XX  
 PA (ZYCO-) ZYCOS INC.  
 XX  
 PI Chicx RM, Tomlinson AJ, Urban RG;  
 XX  
 DR WPI; 2003-040607/03.  
 XX  
 PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 XX  
 PS Example 2; SEQ ID NO 1346; 134pp; English.  
 XX  
 CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 250 AA;  
 XX  
 Query Match 95.6%; Score 1297; DB 6; Length 250;  
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;  
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;  
 XX  
 QY 1 MADNFSIHDALSGSGNPNPGWPGAMGNOPAGAGGYGASYPG-YPGQAPPGAYPGQAPP 59  
 Db 1 MADNFSIHDALSGSGNPNPGWPGAMGNOPAGAGGYGASYPGAYPGQAPPGAYPGQAPP 60  
 QY 60 GAYHGAAGAYPGAYPGAYPGPGPGPGAYPGSSGQPSAPGAY-ANGPYGAPAGPTIIVPNL 118  
 Db 61 GAYHGAAGAYPGAYPGAYPGPGPGPGAYPGSSGQPSATGAYPATGPGAYGAPGAPLIVPNL 120  
 QY 119 PLPGGVVPRMLITLIGTVKPMANRIALDFGNGDVAAHF-PRFENNRRVIVCTKLDNN 177  
 Db 121 PLPGGVVPRMLITLIGTVKPMANRIALDFGNGDVAAHFENRRRVRIVCTKLDNN 180  
 QY 178 WGREERQSVFPFESGKPKIQLVEPDHFXYAVNDAR-LQNHKVKKINISKIGISDI 236  
 Db 181 WGREERQSVFPFESGKPKIQLVEPDHFXYAVNDARHLQYNHKVKKINISKIGISDI 240

QY 237 DLTASATMI 246  
 |||||  
 Db 241 DLTASATMI 250  
 |||||  
 RESULT 23  
 ID ABU04687  
 ID ABU04687 standard; protein; 250 AA.  
 XX  
 AC ABU04687;  
 XX  
 DT 29-JAN-2003 (first entry)  
 XX  
 DE Human expressed protein tag (EPT) #1353.  
 XX  
 XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KM protease; protease inhibitor; transporter; cytoskeletal protein;  
 KM receptor; transcription factor; cancer; MHC;  
 KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200278524-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 28-MAR-2002; 2002MO-US009671.  
 XX  
 XX 28-MAR-2001; 2001US-0279495P.  
 XX 21-MAY-2001; 2001US-0292544P.  
 PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 XX  
 PA (ZYCO-) ZYCOS INC.  
 XX  
 PI Chicx RM, Tomlinson AJ, Urban RG;  
 XX  
 DR WPI; 2003-040607/03.  
 XX  
 PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 XX  
 PS Example 2; SEQ ID NO 1353; 134pp; English.  
 XX  
 CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 250 AA;  
 XX  
 Query Match 95.6%; Score 1297; DB 6; Length 250;  
 Best Local Similarity 97.6%; Pred. No. 4.6e-95;  
 Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

QY 1 MADNFSLHDALSGSGNPNPQGWPGWGMGNQPAAGGYPGASYPG-YFGQAPPGAYPGQAPP 59
DB 1 MADNFSLHDALSGSGNPNPQGWPGWGMGNQPAAGGYPGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHGAAPGAYPGAPAPGVYPPGSPGAYPSSGQPSAPGAY-ATGPGAPAGPLIYVYNL 118
DB 61 GAYPGAPGAYPGAPAPGVYPPGSPGAYPSSGQPSATGAYPATGPGAPAGPLIYVYNL 120
QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFHF-PRNENNRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFHFPRNENNRVIVCNTKLDNN 180
QY 178 WGREERQSVFPFESGKPFKIQLVLEPDHFKVAVNDAH-LQYNHRVKKLNLSKLGISDI 236
DB 181 WGREERQSVFPFESGKPFKIQLVLEPDHFKVAVNDAHLLQYNHRVKKLNLSKLGISDI 240
QY 237 DLTASASYTMI 246
DB 241 DLTASASYTMI 250

```

## RESULT 24

ABU04678 ID ABU04678 standard; protein; 250 AA.

AC ABU04678;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1344.

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC;

KM major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

OS WO200278524-A2.

PN 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0326780P.

PR 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCO8 INC.

XX PA

XX PI Chicx RM, Tomlinson AJ, Urban RG;

XX DR WPI; 2003-040607/03.

XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

XX PT cytoskeletal proteins, receptors or transcription factors), useful for

XX PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or

XX PS leukemia.

XX PS Example 2; SEQ ID NO 1344; 134pp; English.

XX CC The invention describes a purified polypeptide, which comprises a

CC CC fragment of a kinase, phosphatase, protease, protease inhibitor,

CC CC transporter, cytoskeletal protein, receptor or transcription factor. The

CC CC polypeptide is useful as an immunogenic composition for eliciting in a

CC CC mammal an immunogenic response directed against any of the purified

CC CC polypeptide. The purified polypeptide, or the antibody that binds to this

CC CC polypeptide, is useful for treating cancer. The polypeptide is also

CC CC useful for identifying compounds that binds to a naturally processed

CC class I or class II MHC-binding polypeptide. The polypeptides and  
CC polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukaemia. These are also useful for screening agents for  
CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling. Note: This sequence does not appear in the printed  
CC specification but was obtained in electronic format directly from WIP0 at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 250 AA;

QY Query Match 95.6%; Score 1297; DB 6; Length 250;  
DB Best Local Similarity 97.6%; Pred. No. 4, 6e-95;  
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

QY 1 MADNFSLHDALSGSGNPNPQGWPGWGMGNQPAAGGYPGASYPG-YFGQAPPGAYPGQAPP 59
DB 1 MADNFSLHDALSGSGNPNPQGWPGWGMGNQPAAGGYPGASYPGAYPGQAPPGAYPGQAPP 60
QY 60 GAYHGAAPGAYPGAPAPGVYPPGSPGAYPSSGQPSAPGAY-ATGPGAPAGPLIYVYNL 118
DB 61 GAYPGAPGAYPGAPAPGVYPPGSPGAYPSSGQPSATGAYPATGPGAPAGPLIYVYNL 120
QY 119 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFHF-PRNENNRVIVCNTKLDNN 177
DB 121 PLPGGVPRMLITITIGTVKPNANRIALDFQGNDAVFHFPRNENNRVIVCNTKLDNN 180
QY 178 WGREERQSVFPFESGKPFKIQLVLEPDHFKVAVNDAH-LQYNHRVKKLNLSKLGISDI 236
DB 181 WGREERQSVFPFESGKPFKIQLVLEPDHFKVAVNDAHLLQYNHRVKKLNLSKLGISDI 240
QY 237 DLTASASYTMI 246
DB 241 DLTASASYTMI 250

```

## RESULT 25

ABU04679 ID ABU04679 standard; protein; 250 AA.

AC ABU04679;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1345.

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC;

KM major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

OS WO200278524-A2.

PN 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0326780P.

PR 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCO8 INC.

XX PA

XX PI Chicx RM, Tomlinson AJ, Urban RG;

XX DR WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
PT leukemia.  
PS Example 2; SEQ ID NO 1345; 134pp; English.  
XX  
XX The invention describes a purified polypeptide, which comprises a  
CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
CC transporter, cytoskeletal protein, receptor or transcription factor. The  
CC polypeptide is useful as an immunogenic composition for eliciting in a  
CC mammal an immunogenic response directed against any of the purified  
CC polypeptide. The purified polypeptide, or the antibody that binds to this  
CC polypeptide, is useful for treating cancer. The polypeptide is also  
CC useful for identifying compounds that binds to a naturally processed  
CC class I or class II MHC-binding polypeptide. The polypeptides and  
CC polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukemia. These are also useful for screening agents for  
CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling. Note: This sequence does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 250 AA;  
SQ  
Query Match 95.6%; Score 1297; DB 6; Length 250;  
Best Local Similarity 97.6%; Pred. No. 4.6e-95;  
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;  
QY 1 MADNFSIHDLALSGSGNPNPQGWPGAWGNOPAGAGGYPGASYPG-YPGAPPGAYPGQAPP 59  
Db 1 MADNFSIHDLALSGSGNPNPQGWPGAWGNOPAGAGGYPGASYPGAYPGQAPPAYPGQAPP 60  
QY 60 GAYHGAAPGAYPGAPAPGAYVPGPPSGGAYPSSGQSPAPGAY-ATGTYGAPAGGLIYVYNL 118  
Db 61 GAYPGAPGAYPGAPAPGAYVPGPPSGGAYPSSGQSPATGAYPATGTYGAPAGGLIYVYNL 120  
QY 119 PLPGGVVPRMLITLITIGTVKPNANRIALDPQRGNDVAFHF-PRENNRRRYIVCNTKLDNN 177  
Db 121 PLPGGVVPRMLITLITIGTVKPNANRIALDPQRGNDVAFHFPRNNRRRYIVCNTKLDNN 180  
QY 178 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAH-LQYNRHVKKLNEISKLGISGDI 236  
Db 181 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAHLQYNRHVKKLNEISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
Db 241 DLTSASYTMI 250  
RESULT 26  
ABU04675  
ID ABU04675 standard; protein; 250 AA.  
XX  
XX AC ABU04675;  
XX  
XX 29-JAN-2003 (first entry)  
XX  
XX Human expressed protein tag (EPT) #1341.  
XX  
XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
KW protease; protease inhibitor; transporter; cytoskeletal protein;  
KW receptor; transcription factor; cancer; MHC;  
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.  
XX  
XX Homo sapiens.  
XX OS  
XX ID W0200278524-A2.  
XX

PD 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002W0-US009671.  
XX  
XX 28-MAR-2001; 2001US-0279495P.  
XX  
XX 21-MAY-2001; 2001US-0292544P.  
XX  
XX 08-AUG-2001; 2001US-0310801P.  
XX  
XX 01-OCT-2001; 2001US-0326370P.  
XX  
XX 04-DEC-2001; 2001US-0336780P.  
XX  
XX 20-FEB-2002; 2002US-0358985P.  
XX  
XX (ZYCO-) ZYCOs INC.  
XX  
XX Chicz RM, Tomlinson AJ, Urban RG;  
XX  
XX WPI; 2003-040607/03.  
XX  
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
PT leukemia.  
PS Example 2; SEQ ID NO 1341; 134pp; English.  
XX  
XX  
XX The invention describes a purified polypeptide, which comprises a  
CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
CC transporter, cytoskeletal protein, receptor or transcription factor. The  
CC polypeptide is useful as an immunogenic composition for eliciting in a  
CC mammal an immunogenic response directed against any of the purified  
CC polypeptide. The purified polypeptide, or the antibody that binds to this  
CC polypeptide, is useful for treating cancer. The polypeptide is also  
CC useful for identifying compounds that binds to a naturally processed  
CC class I or class II MHC-binding polypeptide. The polypeptides and  
CC polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukemia. These are also useful for screening agents for  
CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling. Note: This sequence does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 250 AA;  
SQ  
Query Match 95.6%; Score 1297; DB 6; Length 250;  
Best Local Similarity 97.6%; Pred. No. 4.6e-95;  
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;  
QY 1 MADNFSIHDLALSGSGNPNPQGWPGAWGNOPAGAGGYPGASYPG-YPGAPPGAYPGQAPP 59  
Db 1 MADNFSIHDLALSGSGNPNPQGWPGAWGNOPAGAGGYPGASYPGAYPGQAPPAYPGQAPP 60  
QY 60 GAYHGAAPGAYPGAPAPGAYVPGPPSGGAYPSSGQSPAPGAY-ATGTYGAPAGGLIYVYNL 118  
Db 61 GAYPGAPGAYPGAPAPGAYVPGPPSGGAYPSSGQSPATGAYPATGTYGAPAGGLIYVYNL 120  
QY 119 PLPGGVVPRMLITLITIGTVKPNANRIALDPQRGNDVAFHF-PRENNRRRYIVCNTKLDNN 177  
Db 121 PLPGGVVPRMLITLITIGTVKPNANRIALDPQRGNDVAFHFPRNNRRRYIVCNTKLDNN 180  
QY 178 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAH-LQYNRHVKKLNEISKLGISGDI 236  
Db 181 WGREERQSVFPFESGKPFKIQLVVEPDHFKVAVNDAAHLQYNRHVKKLNEISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
Db 241 DLTSASYTMI 250  
RESULT 27  
AAG75013  
ID AAG75013 standard; protein; 277 AA.  
XX



QY 1 MADNFSLHDALSGSGNPNPGWPGMNGNOPAGAGYFGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 28 MADNFSLHDALSGSGNPNPGWPGMNGNOPAGAGYFGASYPGQAPPGAYPGQAPP 87  
 QY 60 GAYHGAAPGAYPGAPAPGVYRPGSPGAPYSSSGSPAPGAY-ATGPGYAPAGPLIVRYNL 118  
 DB 88 GAYHGAAPGAYPGAPAPGVYRPGSPGAPYSSSGSPAPGAYPATGPGYAPAGPLIVRYNL 147  
 QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHF-PRFNNRNRYIVCNTKLDNN 177  
 DB 148 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHFPRFNNRNRYIVCNTKLDNN 207  
 QY 178 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAH-LQYNHRYKKLINEISKLIGSDI 236  
 DB 208 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAHLLQYNHRYKKLINEISKLIGSDI 267  
 QY 237 DLTSASYTMI 246  
 DB 268 DLTSASYTMI 277

## RESULT 29

ID ABR56317 standard; protein; 250 AA.  
 ABR56317;

DT 20-NOV-2003 (first entry)

DE Human galectin-3.

KW Human; cytosolic; galectin-3; cancer.

OS Homo sapiens.

Key Location/Qualifiers  
 FT 117..250  
 FT Domain /label=Carbohydrate\_recognition\_domain

PN US2003054982-A1.

PD 20-MAR-2003.

PF 08-JUN-2001; 2001US-00877790.

PR 08-JUN-2001; 2001US-00877790.

PA (JARV/) JARVIS G A.

PA (JOHN/) JOHN C M.

PA (LEFF/) LEFFLER H.

PI Jarvis GA, John CM, Leffler H;

DR WPI; 2003-439728/41.

XX New composition, useful for treating cancer, comprises N-terminally

PT truncated galectin-3 and a carrier.

PS Disclosure; Page 4; 30pp; English.

XX The present invention relates to a composition (I), comprising an amount

CC of N-terminally truncated galectin-3 (ABR56316) and a carrier. The

CC composition is used for treating cancer. The present sequence is the full

CC -length human galectin-3 sequence

XX Sequence 250 AA;  
 QY Query Match 94.8%; Score 1287; DB 6; Length 250;  
 Best Local Similarity 96.8%; Pred. No. 2.9e-94;  
 Matches 242; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

DB 1 MADNFSLHDALSGSGNPNPGWPGMNGNOPAGAGYFGASYPGAYPGQAPPGAYPGQAPP 60  
 QY 60 GAYHGAAPGAYPGAPAPGVYRPGSPGAPYSSSGSPAPGAY-ATGPGYAPAGPLIVRYNL 118  
 DB 61 GAYHGAAPGAYPGAPAPGVYRPGSPGAPYSSSGSPAPGAYPATGPGYAPAGPLIVRYNL 120  
 QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHF-PRFNNRNRYIVCNTKLDNN 177  
 DB 121 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVAFHFPRFNNRNRYIVCNTKLDNN 190  
 QY 178 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAH-LQYNHRYKKLINEISKLIGSDI 236  
 DB 181 WGREEROSVPPFESGKPFKIQVLVEPDHFVAVNDAAHLLQYNHRYKKLINEISKLIGSDI 240  
 QY 237 DLTSASYTMI 246  
 DB 241 DLTSASYTMI 250

## RESULT 30

ID AAR12532 standard; protein; 248 AA.  
 AAR12532;

AC AAR12532;

DT 25-MAR-2003 (revised)

DT 12-SEP-1991 (first entry)

DE Human Macrophage Carbohydrate- and Ige-binding protein.

KW HMEBP; leishmaniasis; Mouse Mac-2; laminin.

OS Homo sapiens.

Key Location/Qualifiers  
 FT 50..55  
 FT Region /note="repeated 5 times"  
 FT 59..64  
 FT Region /note="repeated 5 times"  
 FT 66..71  
 FT Region /note="repeated 5 times"  
 FT 75..80  
 FT Region /note="repeated 5 times"  
 FT 96..101  
 FT Region /note="repeated 5 times"  
 FT 156..161  
 FT Region /note="feature indicated but not described"  
 FT 179..184  
 FT Region /note="feature indicated but not described"

PN MO9108290-A.

PD 13-JUN-1991.

PF 30-NOV-1989; 89US-00444195.

PR 30-NOV-1989; 89US-00444195.

PR 14-SEP-1990; 90US-00582628.

PA (GEHO ) GEN HOSPITAL CORP.

PI Pillal S, Cherayil BJ;

DR WPI; 1991-1919196/26.

DR N-PSDB; AAQ12211.

XX Recombinant gene encoding human macrophage carbohydrate - Ige-binding

PT protein and antibody used to treat, diagnose and prevent e.g.

PT inflammatory bowel disorder, leishmaniasis, hay fever, etc.

PS Claim 1; Fig 3; 36pp; English.

XX This sequence is deduced from a composite of two overlapping cDNA clones

CC isolated from a human colon cancer cell line HT-29 cDNA library. The  
 CC deduced carboxy-terminal half of HMEBP resembles that deduced for mouse  
 CC Mac-2. There is more divergence in the amino-terminal portion. The  
 CC estimated size of this protein is 2x0 smaller than that of its mouse  
 CC counterpart. The two proteins exhibit 77 per cent amino acid sequence  
 CC identity. See also AA012207-012210. (Updated on 25-MAR-2003 to correct PA  
 CC field.)

XX Sequence 248 AA;

Query Match 88.0%; Score 1194; DB 2; Length 248;  
 Best Local Similarity 92.4%; Pred. No. 6.9e-87;  
 Matches 230; Conservative 0; Mismatches 15; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSSGSGNPQGMWPGAMGNQPAAGAGYPPGAYPGQAPPG 60  
 DB 1 MADNFSLHDALSSGSGNPQGMWPGAMGNQPAAGAGYPPGAYPGQAPPG 60  
 QY 61 AYHGAAPGAPGAPAGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119  
 DB 61 AYHGAAPGAPGAPAGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119  
 QY 120 LFGVVPRLITLITGVKPNANRIALDFQRGNDAFHF-PRENENRRVIVCNTKLDNNW 178  
 DB 120 LFGVVPRLITLITGVKPNANRIALDFQRGNDAFHF-PRENENRRVIVCNTKLDNNW 178  
 QY 179 GREERQSVFPESGKPKIQVLEPDHFKVAVNDAH-LQNHVVKKLEISKLGISGID 237  
 DB 180 GREERQSVFPESGKPKIQVLEPDHFKVAVNDAHLLQYHVRVKLEISKLGISGID 239  
 QY 238 LTSASYTMI 246  
 DB 240 LTSASYTMI 248

# RESULT 31

ABU04682  
 ID ABU04682 standard; protein; 248 AA.

AC ABU04682;

DT 29-JAN-2003 (first entry)

XX Human expressed protein tag (BPT) #1348.

XX Translational profiling; expressed protein tag; BPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.

XX Homo sapiens.

XX WO200278524-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US09671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

PA (ZYCO-) ZYCO INC.

PI Chicx RM, Tomlinson AJ, Urban RG;

DR WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.

PS Example 2; SEQ ID NO 1348; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (BPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 248 AA;

Query Match 88.0%; Score 1194; DB 6; Length 248;  
 Best Local Similarity 92.4%; Pred. No. 6.9e-87;  
 Matches 230; Conservative 0; Mismatches 15; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSSGSGNPQGMWPGAMGNQPAAGAGYPPGAYPGQAPPG 60  
 DB 1 MADNFSLHDALSSGSGNPQGMWPGAMGNQPAAGAGYPPGAYPGQAPPG 60  
 QY 61 AYHGAAPGAPGAPAGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119  
 DB 61 AYHGAAPGAPGAPAGVYPPGSGPAGYPPSGQPSAPGAY-ATGPGAPAGPLIVPNLP 119  
 QY 120 LFGVVPRLITLITGVKPNANRIALDFQRGNDAFHF-PRENENRRVIVCNTKLDNNW 178  
 DB 120 LFGVVPRLITLITGVKPNANRIALDFQRGNDAFHF-PRENENRRVIVCNTKLDNNW 178  
 QY 179 GREERQSVFPESGKPKIQVLEPDHFKVAVNDAH-LQNHVVKKLEISKLGISGID 237  
 DB 180 GREERQSVFPESGKPKIQVLEPDHFKVAVNDAHLLQYHVRVKLEISKLGISGID 239  
 QY 238 LTSASYTMI 246  
 DB 240 LTSASYTMI 248

# RESULT 32

AAR12531  
 ID AAR12531 standard; protein; 264 AA.

AC AAR12531;

DT 25-MAR-2003 (revised)

DT 12-SEP-1991 (first entry)

DE Mac2.16 expression product.

XX HMEBP; leishmaniasis; Mouse Mac-2; laminin.

XX Mus musculus.

XX WO9108290-A.

XX 13-JUN-1991.

PF 30-NOV-1989; 89US-00444195.

PR 30-NOV-1989; 89US-00444195.

Query Match	79.7%;	Score 1082;	DB 2;	Length 264;
Best Local Similarity	78.0%;	Pred. No. 5.7e-78;		
Matches 209;	Conservative 16;	Mismatches 17;	Indels 26;	Gaps 8;
QY 1	MADNFSILHATLSSGSGNPNPQGWPGAWGNQ	PAGAGGYPGASYPG-YPGQAPPGAYPGQAP	59	
DB 1	MADSFELNALALGSGNPNPQGYPGAWGNP	-GAGGYPGAAYGAYPGQAPPGAYPGQAP	59	
QY 60	GAYHG-----APGATPGAPAPGYVPPSPSGAYPSSGQSPAPGAY-----		100	
DB 60	GAYPGQAPPSAYVGPPTAPGAYPGPTAPGAYPGQAP	-PGAF--GQPGAYGATPQCSGGYP	116	
QY 101	ATGPGYAPAGGLVPEYNLP	PGGVVPRMLITLIGTVKPNANRIALDPQRGNVAFHF	BR 159	
DB 117	AAGPYGVPAGPLVPIYDLPLPGAVMERMLITMGTVKPNANRIVLDFRRGNVAFHFNDP		176	
QY 160	FNENNRVIVCNTKLDNNMGREERQSVFPESGKPKIQVLVPPDHFKVAVNDAH	-LOYN 218		
DB 177	FNENNRVIVCNTKLDNNMGKEERQSAFPESGKPKRIQVLVADHFKVAVNDAHLLQYN		236	
QY 219	HRVAKLNEISKLGISGDIDLTASAYTI	246		
DB 237	HRMKNLREISQLGISGDITLTLSANHAMT	264		
RESULT 33				
ADC53841				
ID	ADC53841	standard;	protein;	264 AA.
XX	ADC53841;			
XX	ADCS3841;			
XX	18-DEC-2003	(first entry)		
XX	Mouse galectin 3 protein.			
XX	galectin-3; nephritis; glomerular nephritis; antiinflammatory;			
XX	glomerular infiltration; apoptosis; mouse.			
XX	Mus musculus.			
XX	JP2002322082-A.			
XX	08-NOV-2002.			
XX	26-APR-2001;	2001JP--00129200.		
XX	26-APR-2001;	2001JP--00129200.		

XX (PROT-) PROTEGENE KK.  
XX  
XX WPI: 2003-367092/35.  
DR N-PSDB; ADCS3850.  
XX  
XX Agents for prevention and treatment of nephritis, comprise galectin-1,  
PT galectin-3, or galectin-9, by inhibition of intraglomerular infiltration  
PT of leukocytes, CD8 positive cells, and induction of apoptosis of CD8  
PT positive cells.  
XX  
XX Claim 1; SEQ ID NO 2; 31pp; Japanese.  
PS  
XX This invention relates to the use of novel mammal derived galectin-1  
XX (G3), -3 (G3) and -9 (G9) proteins as effective ingredients for  
CC (G3), -3 (G3) and -9 (G9) proteins as effective ingredients for  
CC prevention and treatment of nephritis. The invention discloses agents for  
CC prevention and treatment of nephritis, particularly glomerular nephritis  
CC and may have antiinflammatory activities. The method of the invention  
CC and inhibits glomerular infiltration of leukocytes, CD8 positive cells and  
CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and  
CC apoptosis of CD8 positive cells. The method and sequences of the  
CC invention may be used for prevention and treatment of nephritis,  
CC particularly glomerular nephritis including inhibition of glomerular  
CC infiltration of leukocytes, CD8 positive cells and apoptosis of CD8  
CC positive cells. The present sequence represents the mouse galectin 3  
CC protein of the invention.  
SQ Sequence 264 AA;

[illegible]

```
RESULT 34  
AAR13338  
ID AAR13338 standard; protein; 278 AA.  
XX  
AC  
AA  
X  
AAR13338;  
XX  
DT 25-MAR-2003 (revised)  
PT 12-SEP-1991 (first entry)  
XX  
DE Mac-2 protein including putative signal sequence.  
KW HMEBP; Ieismaniasis; Mouse Mac-2; laminin.  
KX  
XX  
OS Mus musculus.  
  
Key Location/Qualifiers  
FT Peptide 1..14 /label= putative signal peptide  
FT FT /note="Met(1) is encoded by CTG  
FT 15...278  
FT Protein /label= Mac-2
```



XX  
PN WO9108290-A.  
XX  
PD 13-JUN-1991.  
XX  
PF 30-NOV-1989; 89US-00444195.  
XX  
PR 30-NOV-1989; 89US-00444195.  
XX  
PR 14-SEP-1990; 90US-00582628.  
XX  
PA (GENE) GEN HOSPITAL CORP.  
XX  
PI Pillai S, Cherayil BJ;  
XX  
DR WPI: 1991-193196/26.  
XX  
DR N-PSDB; AAQ12210.  
XX  
PT Recombinant gene encoding human macrophage carbohydrate - IgE-binding  
XX  
PT protein and antibody used to treat, diagnose and prevent e.g.  
XX  
PT inflammatory bowel disorder, leishmaniasis, hay fever, etc.  
XX  
PS Disclosure; Fig 2D, 36pp; English.  
XX  
CC This sequence is identical to that deduced from clone Mac 2.16 (see  
XX  
CC AAR12531), except for the putative signal peptide. Mac 2.16 itself does  
XX  
CC not encode a signal peptide, but Mac-2 protein newly synthesised by mouse  
XX  
CC inflammatory peritoneal macrophages was detected in the extracellular  
XX  
CC medium. Clone Mac 2.9 (from which this sequence was deduced) was thus  
XX  
CC analysed for presence of a signal sequence. See also AAQ12207 (for Mac  
XX  
CC 2.16), AAQ12209, AAQ12211. (Updated on 25-MAR-2003 to correct PA field.)  
SQ  
Sequence 278 AA;

Query Match 79.7%; Score 1082; DB 2; Length 278;  
Best Local Similarity 78.0%; Pred. No. 6.1e-78;  
Matches 209; Conservative 16; Mismatches 17; Indels 26; Gaps 8;  
QY 1 MADNPSLHDLASGSGNPNPQGMWPGAGAGGYPGASYPG-YPGQAPPGAYPGQAP 59  
DB 15 MADNPSLHDLASGSGNPNPQGMWPGAGAGGYPGASYPG-YPGQAPPGAYPGQAP 73  
QY 60 GAYHG-----APGAYPGAPAGVYPPGSPGAYSSGQPSAPGAY----- 100  
DB 74 GAYGQAPPGAYPGATAGAPGATAGAPGAYGQAP-PPAAPP--GQPPGAPGAYPGQAP 130  
QY 101 ATGPYGPAGPLIVPNPLPFGGVVPRMLITLGVKNANRILDPORGNDVAFHF-PR 159  
DB 131 AAGYGVVAGPLIVPNPLPFGGVVPRMLITLGVKNANRILDPORGNDVAFHF-PR 190  
QY 160 FENNNRRVIVCNTKLDNNWGEERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAAH-LQYN 218  
DB 191 FENNNRRVIVCNTKLDNNWGEERQSAFPFSGKPFKIQVLVEADHFKVAVNDAAHLLQYN 250  
QY 219 HRVKKLNEISKLGISGIDILTSASYTM 246  
DB 251 HRMKNLREISQIGISGIDILTSANHAM 278

RESULT 35  
AAU97821  
ID AAU97821 standard; protein; 262 AA.  
AC AAU97821;  
XX  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Rat cell membrane anchor protein galectin-3.  
XX  
KW Galectin-3; cell membrane anchor protein; Ras; antisense technology;  
KW farnesylated isoform; H-Ras; K-Ras 4A; K-Ras 4B; N-Ras; mitosis disorder;  
KW cancer; non-malignancy; autoimmune disease; type 1 diabetes; lupus;  
KW multiple sclerosis; cirrhosis; graft rejection; atherosclerosis;  
KW polycystic kidney; post-angioplasty restenosis; cytostatic;

KM Immunosuppressive; antidiabetic; antiatherosclerotic; neuroprotective;  
KM vasotropic; hepatotropic; rat.  
XX  
OS Rattus sp.  
XX  
PN WO200229031-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 01-OCT-2001; 2001WO-IL000918.  
XX  
PR 04-OCT-2000; 2000US-0237858P.  
XX  
PR (UYRA-) UNIV RAMON APPLIED RES & IND DEV LTD.  
XX  
PI Kloog Y, Haklai R, Paz A, El Ad-Sfadia G, Ballan E;  
XX  
DR WPI: 2002-43533/46.  
XX  
PT Identifying anchor proteins that bind Ras protein, by producing complexes  
XX  
PT of Ras and cell membrane proteins in the presence and absence of a Ras  
XX  
PT antagonist and identifying a complex disrupted by the Ras antagonist.  
XX  
PS Disclosure; Page 14, 62pp; English.

The invention describes a method of identifying cell membrane anchor proteins that bind a Ras protein, involving preparing 2 reaction mixtures where one mixture has a Ras antagonist. A cross linking agent is added, and complexes between Ras protein and other proteins are produced. The complexes are then separated and the proteins binding to Ras are identified. The invention also describes a method useful for identifying drug candidates that inhibit aberrant Ras activity. An antisense compound comprising at least one phosphorothioate-modified nucleotide is useful for disrupting aberrant Ras activity in vivo, by infusing the antisense compound into a patient exhibiting this problem. The method is also useful for identifying anchor proteins for the farnesylated isoforms of H-Ras, K-Ras 4A, K-Ras 4B and N-Ras, whose mutated forms are known to be oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful for treating diseases characterised by uncontrolled mitosis, including cancers and various non-malignancies such as autoimmune disease (e.g. type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft rejection, atherosclerosis, polycystic kidneys and post-angioplasty restenosis. This sequence encodes a galectin-3, a cell-membrane anchor protein that binds an isoform of Ras

Query Match 79.5%; Score 1079; DB 5; Length 262;  
Best Local Similarity 77.9%; Pred. No. 9.9e-78;  
Matches 205; Conservative 16; Mismatches 24; Indels 18; Gaps 7;  
QY 1 MADNPSLHDLASGSGNPNPQGMWPGAGAGGYPGASYPG-YPGQAPPGAYPGQAP 59  
DB 1 MADNPSLHDLASGSGNPNPQGMWPGAGAGGYPGASYPG-YPGQAPPGAYPGQAP 59  
QY 60 GAYHG-----APGAYPGAPAGVYPPGSPGAYPSG--GQPSAPGAY-ATGPY 105  
DB 60 SATPDPGPGSAYPGPTAPGATAGAPGATAGAPGATAGAPGATAGAPGATAGAP 119  
QY 106 GATPGLIVPNPLPFGGVVPRMLITLGVKNANRILDPORGNDVAFHF-PRNENN 164  
DB 120 GATPGLIVPNPLPFGGVVPRMLITLGVKNANRILDPORGNDVAFHF-PRNENN 179  
QY 165 RRVIVCNTKLDNNWGEERQSVFPFSGKPFKIQVLVEPDHFKVAVNDAAH-LQYNRVRK 223  
DB 180 RRVIVCNTKLDNNWGEERQSAFPFSGKPFKIQVLVEADHFKVAVNDAAHLLQYNRMKN 239  
QY 224 LNEISKLGISGIDILTSASYTM 246  
DB 240 LNEISKLGISGIDILTSASHAM 262

RESULT 36



```
RESULT 38
ABU04782
ID ABU04782 standard; protein; 264 AA.
XX
AC ABU04782;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human expressed protein tag (EPT) #1448.
XX
KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200278524-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WC-US009671.
XX
PR 28-MAR-2001; 2001US-0279495P.
PR 21-MAY-2001; 2001US-0292544P.
PR 08-AUG-2001; 2001US-0310801P.
PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0355895E.
XX
PA (ZYCO-) ZYCOs INC.
XX
PI Chicx RM, Tomlinson AJ, Urban RG;
XX
DR WPI; 2003-040607/03.
XX
PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX
PS Example 2; SEQ ID NO 1448; 134pp; English.
XX
CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. There are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 264 AA;
Query Match 79.4%; Score 1078; DB 6; Length 264;
Best Local Similarity 78.2%; Pred. No. 1.2e-77;
Matches 208; Conservative 15; Mismatches 21; Indels 22; Gaps 8;
```

```
QY 60 GAYHG--APGAYGAPGAPGVPSPGPGAYPS-----GQSPAGAY-----AT 102
DB 60 GAYPGAPPSATPGCPAPAPAPG-PTAPGAYPGSTAPGAFPGQPGAPGAYPSAPGYPAA 118
QY 103 GPGYAGAGPLIYVYNLPLPGVVPRLITLIGTVKNANRIALDPGRGNDVAFHF-PRFN 161
DB 119 GPGYVAPGFLVYVDLPLPGMLPRLLITIMGTVKNNARIVLDFRGNDVAFHFPRFN 178
QY 162 ENNRRTVNTKLDNNWGREERQSVFPFSGKPFKIQLVLEPDHFKVAVNDAAH-LQYNR 220
DB 179 ENNRRTVNTQDNNWGREERQSAFPFSGKPFKIQLVLEADHFKVAVNDALLQYNRH 238
QY 221 VKLNEISKLGSGDILTSASYTMI 246
DB 239 MGLNRISQLGISGDTLTSANVAMI 264
```

```
RESULT 39
AAR42200
ID AAR42200 standard; protein; 262 AA.
XX
AC AAR42200;
XX
DT 25-MAR-2003 (revised)
DT 18-MAY-1994 (first entry)
XX
DE IGE binding protein.
XX
KW IGE; immunoglobulin; binding protein; allergy; allergic reaction;
KW absorbance; regulation; immune response.
XX
OS Rattus rattus.
XX
PN US5260434-A.
XX
PD 09-NOV-1993.
XX
PF 10-JUL-1991; 91US-00728125.
PR 23-APR-1985; 85US-00726250.
PR 21-SEP-1988; 88US-00247170.
XX
PA (SCRI ) SCRIPPS RES INST.
XX
PI Liu F;
XX
DR WPI; 1993-367951/46.
DR N-PSDB; AAQ50878.
XX
PT DNA encoding IGE-binding protein - with repetitive sequence and homology
PT with IGE receptor.
XX
PS Claim 1; Fig 1; 20pp; English.
XX
CC The Ige binding protein is used to regulate or absorb IGE. It is useful
CC particularly in the absorbance of IGE to ease allergic reactions.
CC (updated on 25-MAR-2003 to correct FF field.)
XX
SQ Sequence 262 AA;
Query Match 79.2%; Score 1075; DB 2; Length 262;
Best Local Similarity 77.6%; Pred. No. 2e-77;
Matches 204; Conservative 17; Mismatches 24; Indels 18; Gaps 7;
```



proteins that bind a Ras protein, involving preparing 2 reaction mixtures where one mixture has a Ras antagonist. A cross linking agent is added, and complexes between Ras protein and other proteins are produced. The complexes are then separated and the proteins binding to Ras are identified. The invention also describes a method useful for identifying drug candidates that inhibit aberrant Ras activity. An antisense compound comprising at least one phosphorothioate-modified nucleotide is useful for disrupting aberrant Ras activity in vivo, by infusing the antisense compound into a patient exhibiting this problem. The method is also useful for identifying anchor proteins for the farnesylated isoforms of H-Ras, K-Ras 4B, K-Ras 4B and N-Ras, whose mutated forms are known to be oncogenic. Reducing or inhibiting aberrant Ras activity in vivo is useful for treating diseases characterised by uncontrolled mitosis, including cancers and various non-malignancies such as autoimmune disease (e.g. type 1 diabetes, lupus and multiple sclerosis), cirrhosis, graft rejection, atherosclerosis, polycystic kidneys and post-angioplasty restenosis. This sequence encodes a galectin-3, a cell-membrane anchor protein that binds an isoform of Ras

Sequence 263 AA;

Query Match 78.4%; Score 1064.5; DB 5; Length 263;  
Best Local Similarity 77.6%; Pred. No. 1.4e-76;  
Matches 208; Conservative 16; Mismatches 17; Indels 27; Gaps 9;

CC 1 MADSLSLNDALAGSGNPNPQGMWAGNQPAGAGYPGAAYPG-YPQAPPGAYPGQAP 59  
DB 1 MADSLSLNDALAGSGNPNPQGMWAGNQPAGAGYPGAAYPG-YPQAPPGAYPGQAP 59  
QY 60 GAYHG-----ARGAYPGAPRGVYRPPSGPGAYPSGQPSAPGAY----- 100  
DB 60 GAYPGQAPPSAYPGPTAPGAYPGPTAPGAYPGQAP-PPGAPF--GQPGAPGAYPPCSCGYP 116  
QY 101 ATGPYAGAPGLIYVYNLPPLPGVVPRLITLITGVKNANRIALDPORGNDVAFHF-PR 159  
DB 117 AAGP-GVYAGPLTVYDPLPGGVWPKLITIMGVKXNANRIYVDFRGNDAVAFHNP 175  
QY 160 FNNNNRRVYVNTKLDNNWGREERQSVFPESGKPKIQVLEPDHFVAVNDAAH-LOYN 218  
DB 176 FNNNNRRVYVNTKLDNNWGREERQSVFPESGKPKIQVLEPDHFVAVNDAAH-LOYN 235  
QY 219 HRVKKNEISKLGISGDIIDLTSAASYTM 246  
DB 236 HRMKNLRISQLGISGDIIDLTSAASYTM 263

RESULT 42

AAW61954  
ID AAW61954 standard; protein; 258 AA.

XX AAW61954;  
AC AAW61954;  
XX 18-SEP-1998 (first entry)  
XX DE Rat galectin amino acid sequence.  
XX KW Mortalin; galectin; diabetes-mediating protein; insulin; DMF; diabetes;  
XX OS drug screening assay.  
XX OS Rattus sp.  
XX PN W09820124-A2.  
XX PD 14-MAY-1998.  
XX PF 24-OCT-1997; 97W0-IB001627.  
XX ER 25-OCT-1996; 96US-0029324P.  
XX ER 05-NOV-1996; 96US-0030088P.  
XX ER 05-NOV-1996; 96US-0030186P.  
XX ER 18-JUL-1997; 97US-00897098.  
XX PA (LARS/) MOSE LARSEN P.

PA (FEYS/) FEY S J.  
PA (NERU/) NERUP J.  
PA (KARL/) KARLSEN A E.  
PA (CHRI/) BIERRE CHRISTENSEN U.  
PA (POCI/) POCTOT F.  
PA (ANDE/) ANDERSEN H U.

PI Mose Larsen P, Fey SJ, Nerup J, Karlseu AE, Bjerre Christensen U;  
PI Poctot F, Andersen HU;  
XX WPI, 1998-286940/25.

XX Identification of diabetes-mediating protein(s) - by transplanting  
XX insulin-secreting cells into host at risk of developing diabetes and  
XX analysing protein expression in transplanted cells.

PS Disclosure; Fig 4; 154pp; English.

XX This represents the amino acid sequence of murine mortalin. This is a  
XX diabetes-mediating protective protein used in the method of invention.  
XX The invention provides methods for in vivo identification of a diabetes-  
XX mediating protein (DMP) by transplanting insulin-secreting cells into  
XX host at risk of developing diabetes and analysing protein expression in  
XX transplanted cells. The DMPs are useful in drug screening assays for  
XX identifying compounds capable of modulating the development of diabetes,  
XX useful as therapeutic agents for the treatment or prevention of diabetes,  
XX and useful as targets of therapeutic agents capable of preventing or  
XX ameliorating diabetes by modulating the expression of the DMP. Changes in  
XX the expression of specific DMPs is diagnostically useful as indicator of  
XX the development of diabetes

Sequence 258 AA;

Query Match 78.1%; Score 1060; DB 2; Length 258;  
Best Local Similarity 78.2%; Pred. No. 3.1e-76;  
Matches 204; Conservative 15; Mismatches 24; Indels 18; Gaps 8;

QY 1 MADSLSLNDALAGSGNPNPQGMWAGNQPAGAGYPGAAYPG-YPQAPPGAYPGQAP 60  
DB 1 MADSLSLNDALAGSGNPNPQGMWAGNQPAGAGYPGAAYPG-YPQAPPGAYPGQAP 59  
QY 61 AYHG-----APGAYPGAPRGVYRPPSGPGAYPSGQPSAPGAY-ATGPYGA 107  
DB 60 AYPGPTGSAVPPTAPGAYPGPTAPGAYPGQAP-PPGAPF--GQPGAPGAYPPCSCGYP 119  
QY 108 PAGPLIYVYNLPPLPGVVPRLITLITGVKNANRIALDPORGNDVAFHF-PRNNRR 166  
DB 120 PTGPLTVYDPLPGGVWPKLITIMGVKXNANRIYVDFRGNDAVAFHNP 178  
QY 167 VIVCNTKLDNNWGREERQSVFPESGKPKIQVLEPDHFVAVNDAAH-LOYNHRVKKLN 225  
DB 179 VIVCNTKLDNNWGREERQSVFPESGKPKIQVLEPDHFVAVNDAAH-LOYNHRVKKLN 237  
QY 226 EISKLGISGDIIDLTSAASYTM 246  
DB 238 EISKLGISGDIIDLTSAASYTM 258

RESULT 43

AAG79833  
ID AAG79833 standard; protein; 143 AA.

XX AAG79833;  
AC AAG79833;  
XX 16-APR-2003 (first entry)  
XX DE N-terminally truncated Galectin-3.  
XX KW Galectin-3; nuclear localisation; BHL domain; Bcl-2; metastasis;  
XX N-terminally truncated; polyethylene glycol; tumour; cancer;  
XX carbohydrate binding; multimerisation; gene therapy.  
XX OS Homo sapiens.



```

XX WPI; 2003-439728/41.
XX
XX New composition, useful for treating cancer, comprises N-terminally
PT truncated galectin-3 and a carrier.
XX
XX Claim 2; Page 4; 30pp; English.
XX
XX The present invention relates to a composition (I), comprising an amount
CC of N-terminally truncated galectin-3 (ABR56316) and a carrier. The
CC composition is used for treating cancer. The present sequence was
CC produced by exhaustive digestion with collagenase
XX
XX Sequence 141 AA;
SQ
Query Match      52.3%; Score 710; DB 6; Length 141;
Best Local Similarity 97.2%; Pred. No. 9.9e-49;
Matches 138; Conservative 1; Mismatches 1; Indels 2; Gaps 2;

QY 106 GAPAGPLIVPYNLPLPGGVPRMLITITIGVKNPNARIALDFORGNDVAFHF-PRFNNENN 164
DB 1 GAPAGPLIVPYNLPLPGGVPRMLITITIGVKNPNARIALDFORGNDVAFHFNNPRFNNENN 59
QY 165 RRVIVCNKLDNNWGRBERQSVFPESGKPEKIQVLVEPDHFVAVNDALH-QYNHRVKKL 224
DB 60 RRVIVCNKLDNNWGRBERQSVFPESGKPEKIQVLVEPDHFVAVNDALH-QYNHRVKKL 119
QY 225 NEISKLGISGIDILTSASYTMI 246
DB 120 NEISKLGISGIDILTSASYTMI 141

RESULT 46
AAP60534
ID AAP60534 standard; protein; 138 AA.
XX
XX AAP60534;
AC
XX 03-OCT-2002 (revised)
DT 01-JAN-1980 (first entry)
XX
XX C-terminal of IGE-binding protein.
DE
XX IGE; immunoglobulin; binding protein; antiallergic; ss.
XX
XX Rattus rattus.
OS
XX WO8606407-A.
XX
XX 06-NOV-1986.
XX
XX 18-APR-1986; 86WO-US000810.
XX
XX 23-APR-1985; 85US-00726250.
XX
XX (MED1-) MEDICAL BIOLOGY INS.
XX
XX Lu FT;
XX
XX WPI; 1986-305131/46.
XX
XX N-ESDB; AAN60456.
XX
XX New IGE-binding polypeptide and its CDNA - may be useful in handling of
PT allergenic reactions.
XX
XX Disclosure; Fig 3; 32pp; English.
XX
XX The IGE-binding protein partially encoded by this C-terminal sequence may
CC be used to establish the structural relatedness of the IGE systems
CC proteins. Specifically, it may be used as an IGE regulator or to absorb
CC IGE in the handling of allergenic reactions. (Updated on 03-OCT-2002 to
CC add missing OS field.)
XX

```

```

SQ Sequence 138 AA;
Query Match      43.7%; Score 593; DB 1; Length 138;
Best Local Similarity 81.9%; Pred. No. 1.9e-39;
Matches 113; Conservative 11; Mismatches 12; Indels 2; Gaps 2;

QY 111 PLIVPYNLPLPGGVPRMLITITIGVKNPNARIALDFORGNDVAFHF-PRFNNNRVIV 169
DB 1 PLIVPYNLPLPGGVPRMLITITIGVKNPNANSITLNFKKGNDIAHFNNPRFNNNRVIV 60
QY 170 CNTKLDNNWGRBERQSVFPESGKPEKIQVLVEPDHFVAVNDALH-QYNHRVKKLDEL 228
DB 61 CNTKLDNNWGRBERQSVFPESGKPEKIQVLVEADHFVAVNDVTL-QYNHRVKKLREIS 120
QY 229 KLGISGIDILTSASYTMI 246
DB 121 QLGITIGDILTSASYTMI 138

RESULT 47
AAB97171
ID AAB97171 standard; protein; 341 AA.
XX
XX AAB97171;
AC
XX 13-AUG-2001 (first entry)
DT
XX
XX Rainbow trout galectin.
DE
XX
XX Rainbow trout; galectin; cytostatic; antiinflammatory; immunity;
XX fish disease; cell adhesion inhibition; inflammation; cancer;
XX T cell movement; apoptosis.
XX
XX Oncorhynchus mykiss.
OS
XX
XX JP2001069976-A.
XX
XX 21-MAR-2001.
XX
XX 01-SEP-1999; 99JP-00247204.
XX
XX 01-SEP-1999; 99JP-00247204.
XX
XX (NORQ ) NORINSUISANSO YOSHOKU KENKYU.
XX
XX WPI; 2001-321173/34.
XX
XX N-PSDB; AAH24641.
XX
XX Novel recombinant Rainbow trout galectin protein and gene encoding the
PT protein, useful for studying fish immunity mechanism, and diagnosis of
PT fish diseases.
XX
XX Claim 1; Page 10-11; 17pp; Japanese.
XX
XX The invention relates to a recombinant protein having a 341 amino acid
CC sequence fully defined in the specification, or its mutant in which at
CC least one amino acid is deleted, substituted or added, but which retains
CC the galectin activity. The protein is useful for studying mechanisms of
CC fish immunity and for diagnosing fish diseases. The galectin proteins are
CC involved in inhibition of cell adhesion, inflammation, metastasis of
CC tumour cells, T cell movement, and apoptosis. The present sequence is the
CC protein of the invention
XX
XX Sequence 341 AA;
SQ
Query Match      23.2%; Score 314.5; DB 4; Length 341;
Best Local Similarity 36.0%; Pred. No. 6.4e-17;
Matches 90; Conservative 30; Mismatches 69; Indels 61; Gaps 17;

QY 9 DALSGSG-----NPNP-----QGWGAMGNQAGAGYFGASYPGYFGAPPGAY 53
DB 132 DTISADGKVELTIVFONPAPTIPADPGFPA---QP---GFP---SYPGFPAQ--PG-F 178

```

```

QY 54 BQAPPGAYHGAAPGAPGVYPPSPGAPSSQSPAPGAYATGPY-GAPAGPL 112
DB 179 P-----SYPGFPA--QPGFSPG-PP--GQPGF-----PYPGFPAQP 211
QY 113 IVPYNIPLPGVGVPMILITLGTVPKPNRRLALDFQRGNDAVAFH-PPRENNRRVIYCN 171
DB 212 AVFYNNMNGIYFPORTINIQGVVNPNNRFFINILFNSGIALHFNPRDEF---LVYEN 268
QY 172 TKLDNNKREERQSVFPFESGKEFKIQVLVEPDHFKVAVNDALQ-VNHRVKKLINEISKL 230
DB 269 SKLRDQWKEERSGGWPFHRGQAFILSLTICDAQCYKIVNGNQTSTYKRRHTLLQOVAIL 328
QY 231 GTSGDILTS 240
DB 329 EVDGDLSTS 338

RESULT 48
ADC53842
ID ADC53842 standard; protein; 322 AA.
XX ADC53842;
AC
XX 18-DEC-2003 (first entry)
DT
XX
DE Mouse galectin 9 protein.
KW galectin-9; nephritis; glomerular nephritis; antinflammatory;
KM glomerular infiltration; apoptosis; mouse.
XX
OS Mus musculus.
XX JP2002322082-A.
PN
XX 08-NOV-2002.
PD
XX 26-APR-2001; 2001JP-00129200.
PF
XX 26-APR-2001; 2001JP-00129200.
PR
XX (PROT-) PROTEGENE KK.
PA
XX WPI: 2003-367092/35.
DR N-PSDB; ADC53851.
XX
XX Agents for prevention and treatment of nephritis, comprise galectin-1,
PT galectin-3, or galectin-9, by inhibition of intraglomerular infiltration
PT of leukocytes, CD8 positive cells, and induction of apoptosis of CD8
PT positive cells.
XX
XX Claim 1; SEQ ID NO 3; 31pp; Japanese.
XX
XX This invention relates to the use of novel mammal derived galectin-1
CC (G1), -3 (G3) and -9 (G9) proteins as effective ingredients for
CC prevention and treatment of nephritis. The invention discloses agents for
CC prevention and treatment of nephritis, particularly glomerular nephritis
CC and may have antinflammatory activities. The method of the invention
CC inhibits glomerular infiltration of leukocytes, CD8 positive cells and
CC apoptosis of CD8 positive cells. The method and sequences of the
CC invention may be used for prevention and treatment of nephritis,
CC particularly glomerular nephritis including inhibition of glomerular
CC infiltration of leukocytes, CD8 positive cells and apoptosis of CD8
CC positive cells. The present sequence represents the mouse galectin 9
CC protein of the invention.
XX
XX Sequence 322 AA;
SQ
Query Match 22.0%; Score 299; DB 7; Length 322;
Best Local Similarity 37.5%; Pred. No. 1e-15;
Matches 69; Conservative 28; Mismatches 47; Indels 40; Gaps 8;
QY 63 HGAPG--AVPGAPAGVYGPSPGAPVSSQSPAPGAYATGPYGAAPAGLIVEYNLF 119

```

```

DB 168 HSTPGQMFSTPGIP-PVYVPTP-----AYT-----IPFYTP 197
QY 120 LPGVVPMILITLGTVPKPNRRLALDFQRGNDAVAFH-PPRENNRRVIYCNPTLDNNW 178
DB 198 IFNGIYPSKSIIMISGNVLPDATRFFHINLRCCGDAFPHNPRENN---AVANNIQINNSW 254
QY 179 GREERQSV--FPFESGKEFKIQVLVEPDHFKVAVNDALH-QYNHRVKKLINEISKLGISGD 235
DB 255 GQERSLLGRPFPSGGSFSTWITCEGHCFKVAVVGQMCERYHHLKXNLQDINTLELVAGD 314
QY 236 IDLT 239
DB 315 IQLT 318

RESULT 49
AAV06997
ID AAV06997 standard; protein; 355 AA.
XX AAV06997;
AC
XX 02-JUL-1999 (first entry)
DT
XX
DE Galectin-9 protein sequence.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM prostate cancer.
XX
OS Homo sapiens.
XX WO904265-A2.
PN
XX 28-JAN-1999.
PD
XX 15-JUL-1998; 98WO-US014679.
PF
XX 17-JUL-1997; 97US-00896164.
PR 10-OCT-1997; 97US-0061599P.
PR 10-OCT-1997; 97US-0061765P.
PR 10-OCT-1997; 97US-00948705.
PR 11-OCT-1997; 97GB-00021697.
PR 22-JUN-1998; 98US-00102322.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
PI O'hare M, Obata Y, Pfeundschn M, Tureci O, Sahin U;
XX
XX WPI: 1999-132448/11.
DR N-PSDB; AAX40198.
XX
XX New isolated cancer associated nucleic acids and polypeptides - isolated
PT using sera from cancer patients, used to develop products for the
PT diagnosis, monitoring or treatment of cancers.
XX
XX Example 8; Page 779-780; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer
XX

```



SQ Sequence 355 AA;

Query Match 21.3%; Score 289; DB 2; Length 355;  
Best Local Similarity 36.2%; Pred. No. 7e-15;  
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;

QY 66 PGAYGAPAAAA-----PGVYPPSPG-----PGAYPSSGPP-----SAPGAYAT 102  
DB 150 PRTVPQPAFSTVPFSQPCPPPRGRGRQKPPGVWPANPAPITQTVIHVQSAFGQWFS 209  
QY 103 GP-----YGAPAGPLIVPNLPLPGVVPRLITITGTVKPNANRLADFGQNDVAFH 156  
DB 210 TPAIPPMVYPHPAYPM--PFTITLIGLYPSKSTILSGTVLPSAQRFHINLCSGNHIAFH 267  
QY 157 F-PFENNRRVIVCNTKLDNNMGREERQ--SVPFESGKPKIQLVLEPDHFKVAVDA 213  
DB 268 LNPFFDEN---AVVRNTQIDNSWGSEERSLPRKMPFVRGGSFVWILCEAHCLKVAVDGQ 324  
QY 214 HL-QYNHRVKKLNEISKIGISGIDILT 239  
DB 325 HLFEPYHRLRLPTINRLLEVGGDIQLT 351

## RESULT 50

AAW85664

ID AAW85664 standard; protein; 355 AA.

XX AAW85664;

XX 19-JUL-1999 (first entry)

XX Galectin-9 like protein.

XX Galectin-9; lectin; galactose; Hodgkin's disease; pharmaceutical;

XX sugar chain; intercellular adhesion; cell proliferation.

XX Homo sapiens.

XX W09910490-A1.

XX 04-MAR-1999.

XX 19-AUG-1998; 98WO-JP003670.

XX 22-AUG-1997; 97JP-00226468.

XX (SAGA ) SAGAMI CHEM RES CENTRE.

XX (PROT-) PROTEGENE INC.

XX Kato S, Yamaguchi T, Sekine S, Kamata K;

XX WPI; 1999-228913/19.

XX N-PSDB; AAX08490, AAX08491.

XX A new human protein having lactose binding properties.

XX Claim 2; Page 55-57; 64pp; English.

CC Galectins are the general term for animal lectins binding to galactose.  
CC Animal lectins exist in many sites such as the cytoplasm, the nucleus,  
CC the cell membrane etc. and are considered to be associated with cell  
CC proliferation. Galectin-9 has been identified as an antigenic protein  
CC reacting with an antibody contained in the serum of Hodgkin's disease and  
CC has a structure where two sugar chain-binding domains are connected by a  
CC linker peptide. The true role of galectin-9 in the body has not been  
CC completely identified but is thought to be involved in intercellular  
CC adhesion. The human galectin-9 like protein coding sequences are  
CC characterised by containing the sequence described in AAX08489. The  
CC protein can be used as pharmaceuticals or reagents for sugar chain  
CC research. The cDNA is used as a probe for gene diagnosis and for gene  
CC therapy  
CC  
SQ Sequence 355 AA;

Query Match 21.3%; Score 289; DB 2; Length 355;  
Best Local Similarity 36.2%; Pred. No. 7e-15;  
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;

QY 66 PGAYGAPAAAA-----PGVYPPSPG-----PGAYPSSGPP-----SAPGAYAT 102  
DB 150 PRTVPQPAFSTVPFSQPCPPPRGRGRQKPPGVWPANPAPITQTVIHVQSAFGQWFS 209  
QY 103 GP-----YGAPAGPLIVPNLPLPGVVPRLITITGTVKPNANRLADFGQNDVAFH 156  
DB 210 TPAIPPMVYPHPAYPM--PFTITLIGLYPSKSTILSGTVLPSAQRFHINLCSGNHIAFH 267  
QY 157 F-PFENNRRVIVCNTKLDNNMGREERQ--SVPFESGKPKIQLVLEPDHFKVAVDA 213  
DB 268 LNPFFDEN---AVVRNTQIDNSWGSEERSLPRKMPFVRGGSFVWILCEAHCLKVAVDGQ 324  
QY 214 HL-QYNHRVKKLNEISKIGISGIDILT 239  
DB 325 HLFEPYHRLRLPTINRLLEVGGDIQLT 351

Search completed: August 23, 2004, 14:37:21  
Job time : 69 secs

**This Page Blank (uspto)**

Gencore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:38:48 ; Search time 123 Seconds  
(without alignments)  
628.510 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357  
Sequence: 1 MADNPSLHDLALSGSGNPNPQ.....ISKLGISGIDITLSASYTWI 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database :  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/ECT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/ECTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US09A\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10C\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1315	96.9	250	9	US-09-263-689-10
2	1315	96.9	250	9	US-09-981-353-127
3	1315	96.9	250	10	US-09-877-790-2
4	1315	96.9	250	10	US-09-919-039-298
5	1315	96.9	250	12	US-10-133-234A-1
6	1315	96.9	250	12	US-10-133-234A-6
7	1315	96.9	250	14	US-10-235-674-11
8	1315	96.9	250	14	US-10-235-674-10
9	1315	96.9	250	16	US-10-153-668-346
10	1305	96.2	250	12	US-10-398-519-10
11	1297	95.6	250	12	US-10-398-519-6
12	1297	95.6	250	12	US-10-398-519-8
13	1297	95.6	250	12	US-10-398-519-12
14	1297	95.6	277	14	US-10-106-698-5787
15	1079	79.5	262	9	US-09-263-689-14

16	1079	79.5	262	12	US-10-398-519-13	Sequence 13, Appl
17	1079	79.5	262	14	US-10-235-674-14	Sequence 14, Appl
18	1064.5	77.3	263	12	US-10-398-519-14	Sequence 14, Appl
19	1049.5	77.3	245	12	US-10-133-234A-6	Sequence 6, Appl
20	1024	75.5	242	12	US-10-133-234A-4	Sequence 4, Appl
21	752	55.4	149	9	US-09-877-790-6	Sequence 6, Appl
22	723	53.3	143	10	US-10-409-786-1	Sequence 1, Appl
23	723	53.3	143	16	US-10-409-786-2	Sequence 1, Appl
24	723	53.3	144	16	US-10-409-786-2	Sequence 2, Appl
25	695.5	51.3	262	12	US-10-133-234A-5	Sequence 5, Appl
26	299	22.0	353	12	US-10-633-035-8	Sequence 8, Appl
27	295	21.7	322	9	US-09-728-479-11	Sequence 11, Appl
28	289	21.3	355	12	US-10-415-566-1	Sequence 1, Appl
29	289	21.3	355	12	US-10-633-035-2	Sequence 2, Appl
30	289	21.3	355	12	US-10-633-035-6	Sequence 6, Appl
31	289	21.3	355	14	US-10-376-133-18	Sequence 18, Appl
32	289	21.3	378	9	US-09-738-973-439	Sequence 439, App
33	289	21.3	378	9	US-09-854-133-439	Sequence 439, App
34	289	21.3	378	14	US-10-144-649A-439	Sequence 439, App
35	288	21.2	139	12	US-10-133-234A-3	Sequence 3, Appl
36	283	20.9	323	9	US-09-728-479-2	Sequence 2, Appl
37	283	20.9	323	12	US-10-415-566-2	Sequence 2, Appl
38	283	20.9	323	14	US-10-024-258A-141	Sequence 14, App
39	283	20.9	323	14	US-10-042-211A-141	Sequence 14, App
40	283	20.9	323	16	US-10-617-217A-141	Sequence 14, App
41	271	20.0	355	15	US-10-138-588-76	Sequence 76, Appl
42	267	19.7	322	12	US-10-633-035-7	Sequence 7, Appl
43	267	19.7	323	9	US-09-728-479-12	Sequence 12, Appl
44	264	19.5	311	9	US-09-263-689-4	Sequence 4, Appl
45	264	19.5	311	12	US-10-415-566-3	Sequence 3, Appl
46	264	19.5	311	14	US-10-235-674-8	Sequence 8, Appl
47	250.5	18.5	145	9	US-09-728-479-8	Sequence 8, Appl
48	250.5	18.5	145	9	US-09-894-526-5	Sequence 5, Appl
49	250.5	18.5	145	9	US-09-894-526-5	Sequence 5, Appl
50	250.5	18.5	145	14	US-10-235-674-12	Sequence 12, Appl
51	250.5	18.5	145	14	US-10-156-136-42	Sequence 42, Appl
52	250.5	18.5	145	15	US-10-455-366-3	Sequence 3, Appl
53	249	18.3	149	9	US-09-894-526-3	Sequence 3, Appl
54	246.5	18.2	168	9	US-09-922-217-139	Sequence 199, App
55	246.5	18.2	168	9	US-09-833-263-199	Sequence 199, App
56	246.5	18.2	168	13	US-10-025-380-199	Sequence 1, Appl
57	243.5	17.9	145	9	US-09-894-526-1	Sequence 1, Appl
58	240.5	17.7	246	15	US-10-138-588-78	Sequence 78, Appl
59	236.5	17.4	324	9	US-09-728-479-7	Sequence 7, Appl
60	236.5	17.4	324	9	US-09-263-689-11	Sequence 11, Appl
61	236.5	17.4	324	14	US-10-235-674-11	Sequence 11, Appl
62	228.5	16.8	359	14	US-10-376-133-20	Sequence 20, Appl
63	225.5	16.6	351	14	US-10-106-698-4502	Sequence 4502, Ap
64	223.5	16.5	183	10	US-09-948-227-8	Sequence 8, Appl
65	223.5	16.5	200	9	US-09-263-689-8	Sequence 8, Appl

# ALIGNMENTS

RESULT 1  
US-09-263-689-10  
Sequence 10, Application US/09263689  
Patent No. US20020150970A1  
GENERAL INFORMATION:  
APPLICANT: NI, Ulan  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Steine, Kessler, Goldstein, & Fox P.L.L.C.  
STREET: 1100 New York Ave., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/263,689  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/946,914  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Steffe, Eric K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 250 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-263-689-10

Query Match 96.9%; Score 1315; DB 9; Length 250;  
Best Local Similarity 98.4%; Pred. No. 9.8e-87;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNPNPQGWPGAMGNQPAAGGYPGASYPG-YPGQAPPGAYPGQAPP 59  
DB 1 MANNFSIHDALSGSGNPNPQGWPGAMGNQPAAGGYPGASYPGAYPGQAPP 60  
QY 60 GAYHGAAPGAYPAPGAYVPGPPSGPAGYSSGQPSAPGAY-ATGPGAGAPGLIIVPNYL 118  
DB 61 GAYHGAAPGAYPAPGAYVPGPPSGPAGYSSGQPSAPGAYPATGPGAGAPGLIIVPNYL 120  
QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVAFH-PRFNNRRRVIYVNTKLDNN 177  
DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVAFHPRFNNRRRVIYVNTKLDNN 180  
QY 178 WGRERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LOYNHRVKKLNEISKLGISGDI 236  
DB 181 WGRERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLQYNHRVKKLNEISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 2  
US-09-981-353-127  
Sequence 127; Application US/09981353  
Patent No. US20020160382A1  
GENERAL INFORMATION:  
APPLICANT: Lasek, Amy W.  
APPLICANT: Jones, David A.  
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
FILE REFERENCE: PA-0038 US  
CURRENT APPLICATION NUMBER: US/09/981,353  
CURRENT FILING DATE: 2001-10-11  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: PERL Program  
SEQ ID NO 127  
LENGTH: 250  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20020160382A1 898779CDD1  
US-09-981-353-127

Query Match 96.9%; Score 1315; DB 9; Length 250;  
Best Local Similarity 98.4%; Pred. No. 9.8e-87;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNPNPQGWPGAMGNQPAAGGYPGASYPG-YPGQAPPGAYPGQAPP 59  
DB 1 MANNFSIHDALSGSGNPNPQGWPGAMGNQPAAGGYPGASYPGAYPGQAPP 60  
QY 60 GAYHGAAPGAYPAPGAYVPGPPSGPAGYSSGQPSAPGAY-ATGPGAGAPGLIIVPNYL 118  
DB 61 GAYHGAAPGAYPAPGAYVPGPPSGPAGYSSGQPSAPGAYPATGPGAGAPGLIIVPNYL 120  
QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVAFH-PRFNNRRRVIYVNTKLDNN 177  
DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVAFHPRFNNRRRVIYVNTKLDNN 180  
QY 178 WGRERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LOYNHRVKKLNEISKLGISGDI 236  
DB 181 WGRERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLQYNHRVKKLNEISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 3  
US-09-877-790-2  
Sequence 2; Application US/09877790  
Publication No. US20030054982A1  
GENERAL INFORMATION:  
APPLICANT: Jarvis, Gary  
APPLICANT: John, Constance  
APPLICANT: Lefler, Hakon  
TITLE OF INVENTION: N-TERMINALLY TRUNCATED GALECTIN-3 FOR USE IN TREATING CANCER  
FILE REFERENCE: 3157.00004  
CURRENT APPLICATION NUMBER: US/09/877,790  
CURRENT FILING DATE: 2001-06-08  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 250  
TYPE: PRT  
ORGANISM: homo sapien  
US-09-877-790-2

Query Match 96.9%; Score 1315; DB 10; Length 250;  
Best Local Similarity 98.4%; Pred. No. 9.8e-87;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSIHDALSGSGNPNPQGWPGAMGNQPAAGGYPGASYPG-YPGQAPPGAYPGQAPP 59  
DB 1 MANNFSIHDALSGSGNPNPQGWPGAMGNQPAAGGYPGASYPGAYPGQAPP 60  
QY 60 GAYHGAAPGAYPAPGAYVPGPPSGPAGYSSGQPSAPGAY-ATGPGAGAPGLIIVPNYL 118  
DB 61 GAYHGAAPGAYPAPGAYVPGPPSGPAGYSSGQPSAPGAYPATGPGAGAPGLIIVPNYL 120  
QY 119 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVAFH-PRFNNRRRVIYVNTKLDNN 177  
DB 121 PLPGGVVPRMLITLITGVKPNANRIALDFQGNDAVAFHPRFNNRRRVIYVNTKLDNN 180  
QY 178 WGRERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAH-LOYNHRVKKLNEISKLGISGDI 236  
DB 181 WGRERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLQYNHRVKKLNEISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 4  
US-09-919-039-298

; Sequence 296, Application US/09919039  
; Publication No. US2003010871A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
; FILE REFERENCE: PA-0035 US  
; CURRENT APPLICATION NUMBER: US/09/919, 039  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: 60/222,113  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 401  
; SOFTWARE: PERL Program  
; SEQ ID NO 298  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030108871A1 898779GCD1  
US-09-919-039-298

Query Match 96.9%; Score 1315; DB 10; Length 250;  
Best Local Similarity 98.4%; Pred. No. 9, 8e-87;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPG-YPGQAPPGAYPGQAP 59  
DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPGAYPGQAPPGAYPGQAP 60  
QY 60 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPSSGQPSAPGAY-ATGPGYGAAPGAPLIVPNYL 118  
DB 61 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPSSGQPSAPGAYPATGPGYGAAPGAPLIVPNYL 120  
QY 119 PLPGGVPRMLITLITIGTVKPNANRIALDFQGNDAVAFHF-PPFNENNRVIVCNTKLDNN 177  
DB 121 PLPGGVPRMLITLITIGTVKPNANRIALDFQGNDAVAFHFPPFNENNRVIVCNTKLDNN 180  
QY 178 WGERERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAH-IQYNHRVKKLNESKLGISGDI 236  
DB 181 WGERERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAHILQYNHRVKKLNESKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 5  
US-10-133-234A-1  
; Sequence 1, Application US/10133234A  
; Publication No. US20040071684A1  
; GENERAL INFORMATION:  
; APPLICANT: Panjwani et al.  
; TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the  
; FILE REFERENCE: 2002458-0006  
; CURRENT APPLICATION NUMBER: US/10/133,234A  
; CURRENT FILING DATE: 2002-04-26  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Human  
US-10-133-234A-1

Query Match 96.9%; Score 1315; DB 12; Length 250;  
Best Local Similarity 98.4%; Pred. No. 9, 8e-87;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPG-YPGQAPPGAYPGQAP 59  
DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPGAYPGQAPPGAYPGQAP 60

QY 60 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPSSGQPSAPGAY-ATGPGYGAAPGAPLIVPNYL 118  
DB 61 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPSSGQPSAPGAYPATGPGYGAAPGAPLIVPNYL 120  
QY 119 PLPGGVPRMLITLITIGTVKPNANRIALDFQGNDAVAFHF-PPFNENNRVIVCNTKLDNN 177  
DB 121 PLPGGVPRMLITLITIGTVKPNANRIALDFQGNDAVAFHFPPFNENNRVIVCNTKLDNN 180  
QY 178 WGERERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAH-IQYNHRVKKLNESKLGISGDI 236  
DB 181 WGERERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAHILQYNHRVKKLNESKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 6  
US-10-398-519-11  
; Sequence 11, Application US/10398519  
; Publication No. US20040072258A1  
; GENERAL INFORMATION:  
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &  
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.  
; TITLE OF INVENTION: ISOPRENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS  
; FILE REFERENCE: THREEOS 3.4-007  
; CURRENT APPLICATION NUMBER: US/10/398,519  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: 60/237,858  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-398-519-11

Query Match 96.9%; Score 1315; DB 12; Length 250;  
Best Local Similarity 98.4%; Pred. No. 9, 8e-87;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPG-YPGQAPPGAYPGQAP 59  
DB 1 MADNFSLHDALSGSGNPNPQGMWPGAMGNQAPAGAGYPGASYPGAYPGQAPPGAYPGQAP 60  
QY 60 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPSSGQPSAPGAY-ATGPGYGAAPGAPLIVPNYL 118  
DB 61 GAYHGAAPGAYPGAPAPGAYPGPPSGGAYPSSGQPSAPGAYPATGPGYGAAPGAPLIVPNYL 120  
QY 119 PLPGGVPRMLITLITIGTVKPNANRIALDFQGNDAVAFHF-PPFNENNRVIVCNTKLDNN 177  
DB 121 PLPGGVPRMLITLITIGTVKPNANRIALDFQGNDAVAFHFPPFNENNRVIVCNTKLDNN 180  
QY 178 WGERERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAH-IQYNHRVKKLNESKLGISGDI 236  
DB 181 WGERERQSVFPFESGKPEKIQVLVEPDHFKVAVNDAHILQYNHRVKKLNESKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 7  
US-10-235-674-10  
; Sequence 10, Application US/10235674  
; Publication No. US20030040081A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides  
; FILE REFERENCE: 1488.0560004



```

Db      121 PLPGGVPRMLITLIGTVKPNANRIALDFORGNDVAFHFHNPFRNNRRVIVCNTKLDNN 180
;
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-6
Query Match      95.6%; Score 1297; DB 12; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.9e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

```

RESULT 10
US-10-398-519-10
; Sequence 10, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 10-04
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-10

```

```

Query Match      96.2%; Score 1305; DB 12; Length 250;
Best Local Similarity 98.0%; Pred. No. 5.1e-86;
Matches 245; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

```

```

;
; US-10-398-519-6
; Sequence 6, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6

```

```

;
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-6

```

```

Query Match      95.6%; Score 1297; DB 12; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.9e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

```

;
; US-10-398-519-8
; Sequence 8, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-8

```

```

;
; US-10-398-519-8
; Sequence 8, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-398-519-8

```

```

Query Match      95.6%; Score 1297; DB 12; Length 250;
Best Local Similarity 97.6%; Pred. No. 1.9e-85;
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

```

;
; US-10-398-519-6
; Sequence 6, Application US/10398519
; Publication No. US20040072258A1
; GENERAL INFORMATION:
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
; TITLE OF INVENTION: ISOPENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
; FILE REFERENCE: THYREOS 3.4-007
; CURRENT APPLICATION NUMBER: US/10/398,519
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/237,858
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6

```

Db 241 DLTSASYMTI 250

RESULT 13  
US-10-398-519-12  
; Sequence 12, Application US/10398519  
; Publication No. US20040072258A1  
; GENERAL INFORMATION:  
; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &  
; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.  
; TITLE OF INVENTION: ISOPRENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS  
; FILE REFERENCE: THYEOS 3.4-007  
; CURRENT APPLICATION NUMBER: US/10/398,519  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: 60/237,858  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-398-519-12

Query Match 95.6%; Score 1297; DB 12; Length 250;  
Best Local Similarity 97.6%; Pred. No. 1.9e-85;  
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSNPNPQGMFGAMGNOAGAGGYPGASYPG-YPGQAPPGAYPGQAP 59  
Db 1 MADNFSLHDALSGSNPNPQGMFGAMGNOAGAGGYPGASYPGAYPGQAPPGAYPGQAP 60  
QY 60 GAYHGAAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAY-ATGPGAPAGPLIVPYNL 118  
Db 61 GAYPGAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAYPATGPGYGAAPAGPLIVPYNL 120  
QY 119 PLPGGVPERMLTTITIGTKVPMNARIALDPQRGNDVAEHF-PRFNENNRVIVCNTKLDNN 177  
Db 121 PLPGGVPERMLTTITIGTKVPMNARIALDPQRGNDVAEHFPRFNENNRVIVCNTKLDNN 180  
QY 178 WGEREROSVPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNHRVKKLEISKLGISGDI 236  
Db 181 WGEREROSVPFESGKPFKIQVLVEPDHFKVAVNDAHLQYNHRVKKLEISKLGISGDI 240  
QY 237 DLTSASYMTI 246  
Db 241 DLTSASYMTI 250

RESULT 14  
US-10-106-698-5787  
; Sequence 5787, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,117  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: Patentin Ver. 3.0  
; SEQ ID NO 5787  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE

LOCATION: (12)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (91)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (125)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-106-698-5787

Query Match 95.6%; Score 1297; DB 14; Length 277;  
Best Local Similarity 97.6%; Pred. No. 2.1e-85;  
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSNPNPQGMFGAMGNOAGAGGYPGASYPG-YPGQAPPGAYPGQAP 59  
Db 28 MADNFSLHDALSGSNPNPQGMFGAMGNOAGAGGYPGASYPGAYPGQAPPGAYPGQAP 87  
QY 60 GAYHGAAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAY-ATGPGAPAGPLIVPYNL 118  
Db 88 GAYXGAPGAYPGAPAPGAYPGPSPGAYPSSGQPSAPGAYPATGPGYGAAPAGPLIVPYNL 147  
QY 119 PLPGGVPERMLTTITIGTKVPMNARIALDPQRGNDVAEHF-PRFNENNRVIVCNTKLDNN 177  
Db 148 PLPGGVPERMLTTITIGTKVPMNARIALDPQRGNDVAEHFPRFNENNRVIVCNTKLDNN 207  
QY 178 WGEREROSVPFESGKPFKIQVLVEPDHFKVAVNDAH-LQYNHRVKKLEISKLGISGDI 236  
Db 208 WGEREROSVPFESGKPFKIQVLVEPDHFKVAVNDAHLQYNHRVKKLEISKLGISGDI 267  
QY 237 DLTSASYMTI 246  
Db 268 DLTSASYMTI 277

RESULT 15  
US-09-263-689-14  
; Sequence 14, Application US/09263689  
; Patent No. US20020150970A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/263,689  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/946,914  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K. 688  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488, 0560001/EKS/SGM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:



```

; LENGTH: 262 amino acids
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-263-689-14

```

Query Match	79.5%;	Score 1079;	DB 9;	Length 262;
Best Local Similarity	77.9%;	Pred. No. 8.9e-70;		
Matches 205				

QY 60 GATHG-----AGCAYGADAPGVYCPGPGSGAYPSS--GGPSAPGAY-ATGPE 105

Db 60 SAPPGETGPSAPYGPPTAPAPYGPPTAPAPAFPGQPGGPGAPYSPAGAYSPAPATGPF 119

QY 106 GAPAGELIPEYNIPLPGGVVPRMLLTIIIGTKPKNANRIALDFQGNVVAHHF-PRERNEN 164

Db 120 GAPGTGLIYPIYMPPLPGGVMPRLTLITIGYKPNANSTLTINEKKGANDIAEHFPRNENK 179

QY 165 RRVIVCNTRYLDNNMGREEROSVFPPESGKPEPKIVLVEPDHFKAIVADAH-LQYNRVKK 223

Db 180 RRVIVCNTRYKQDNNMGREEROSAFPPESGKPEPKIVLVEADHFKAIVADVALLQYNRMK 239

QY 224 LNEISLGISSGIDITLSASYTKI 246

Db 240 LBREISQGIIGDITLTLSASHAMT 262

RESULT 16  
US-10-398-519-13

```

1  APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &
2  APPLICANT: INDUSTRIAL DEVELOPMENT LTD.
3  TITLE OF INVENTION: ISORENIN-D-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS
4  FILE REFERENCE: TH9805 3.4-007
5  CURRENT APPLICATION NUMBER: US/10/398,519
6  CURRENT FILING DATE: 2003-04-04
7  PRIOR APPLICATION NUMBER: 60/237,858
8  PRIOR FILING DATE: 2000-10-04
9  NUMBER OF SEQ ID NOS: 33
10 SOFTWARE: PatentIn Ver. 2.1
11 SEQ ID NO 13
12     LENGTH: 262
13     TYPE: PRT
14 ORGANISM: Rattus sp.
15 OS-10-398-519-13

```

Query Match	79.5%;	Score 1079;	DB 12;	Length 262
Best [local] similarity	77.9%;	Prod No 9 00 70		

QY 1MADNLSLHDALSSGSGNPEDQCPGAMGNGQZAGAGYPCGASVPG--YPCQAPPGAYPGQAPP 59  
 QY 1MADGSLINDALGSGANPQCPGMGAMQGP--GAGGYPCASYPGAYPCQAPPGPGQAPP 59  
 Db 1  
 QY 60 GATHG-----AGAYPGAPAPGVDPGPGSGGAYSS--GQSPASGAY-ATGPY 105  
 Db 60 SAYPGPTGSAVPGTAPTAGAYPGTAPGAPFGQGGGFGAYPSANGATPSPAGAYPATGPF 119  
 QY 106 GAPAPELLVYNLPLPGGVVPRMLITLIGYTKEPVANKIALDPQGNDAVHF--PRENNN 164  
 Db 120 GAPTPELVVPYDMLPGGVMPRMILITLIGYTKEPVANSILTFKKGNDAIACHFNDRFENN 179  
 QY 165 RRVIYCNKTLNNMREERQGVPPFESGSKPKIQLVLEDPHFKAIVANDAA--LQYNNRVKK 223  
 Db 180 RRVIYCNKQDNNMREERQSAFFPESGSKPKIQLVLEADHFKAIVANDVALLQYNNRMKN 239

```

QY      224 LNEISKLGISGDIDLTSASYTMI 246
          |||:|||||: ||
Db      240 LREISQLGIGDITLTSASHAMI 262

```

RESULT 17  
US-10-235-674-14  
; Sequence 14, Application US/10235674  
; Publication No. US20030040081A1  
GENERAL INFORMATION

```

?      APPLICANT: Nü, Jian
?      APPLICANT: Gentz, Reiner L.
?      APPLICANT: Ruben, Steven M.
?      TITLE OF INVENTION: Galactin 9 and 10SV Polynucleotides
?      FILE REFERENCE: 1488.0560004
?      CURRENT APPLICATION NUMBER: US/10/235,674
?      CURRENT FILING DATE: 2002-09-06
?      PRIOR APPLICATION NUMBER: US 09/656,450
?      PRIOR FILING DATE: 2000-09-06
?      PRIOR APPLICATION NUMBER: US 09/263,689
?      PRIOR FILING DATE: 1999-03-05
?      PRIOR APPLICATION NUMBER: US 08/946,914
?      PRIOR FILING DATE: 1997-10-09
?      PRIOR APPLICATION NUMBER: US 60/028,093
?      PRIOR FILING DATE: 1996-10-09
?      NUMBER OF SEQ ID NOS: 60
?      SOFTWARE: PatentIn version 3.0
?      SEQ ID NO 14
?      LENGTH: 262
?      TYPE: PRT
?      ORGANISM: Rat
US-10-235-674-14

```

Query Match	79.5%;	Score 1079;	DB 14;	Length 262;
Best Local Similarity	77.9%;	Pred. No. 8.9e-70;		
Matches 205; Conservative	16;	Mismatches 24;	Indels 18;	Gaps 7;

QY 1MADNLSLHMLASGSSGNPNPDGCPGAGMGQCPAGAGYPGASYPG-YPGQAPPGAYPGQAP 59  
 Db 1MADGSLINDALAGSGNPNPDGCPGAGMGQCPGAGYPGASYPGAYPGQAPPGGYPGQAP 59  
 QY 60 GAHYG-----APGAYPGAPAPGVDPGPPSGGAYPSS--GGPSAPGAY-ATGPY 105  
 Db 60 SAYPPTGTSAYPGPTAGAYPGPTAPGAPFGCGGCGGAGYPSAPGAYPSAPGAYPATGPF 119  
 QY 106 GAPAPGLVYPNYLPLPGGVVPRLMTITLTGTVKPNVNLALDFQGNVVAEHH--PRFENN 164  
 Db 120 GAPTPGLVYPDMPLPGGVMPRLMTITLTGTVKPNANSTITLNFKGNGLIAHFFRERNENN 179  
 QY 165 RRVIVCNTKLDNNNGREERQGVFPFESGKPEKIVLVEPDHFYKAVNDAA-LQYNHVKK 223  
 Db 180 RRVIVCNTKQNNMGREERQSGAPPFESGKPEKIVLVEADHFYKAVNDVHLLQYNHMKM 239  
 QY 224 INEISKLIGSDILTSASYTMI 246  
 Db 240 IREISQLIGIDITLTSSASHMI 262

RESULT 18  
 US-10-398-519-14  
 ; Sequence 14, Application US/10398519  
 ; Publication No. US20040072258A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &  
 ; APPLICANT: INDUSTRIAL DEVELOPMENT LTD.  
 ; TITLE OF INVENTION: ISOPRENOID-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS  
 ; FILE REFERENCE: THREEOS 3.4-007  
 ; CURRENT APPLICATION NUMBER: US/10/398,519  
 ; CURRENT FILING DATE: 2003-04-04  
 ; PRIOR APPLICATION NUMBER: 60/227,858  
 ; PRIOR FILING DATE: 2000-10-04  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.1

APPLICANT: RAMOT UNIVERSITY AUTHORITY FOR APPLIED RESEARCH &  
 APPLICANT: INDUSTRIAL DEVELOPMENT LTD.  
 TITLE OF INVENTION: ISOGENOIC-DEPENDENT RAS ANCHORAGE (IDRA) PROTEINS  
 FILE REFERENCE: THYROS 3.4-007  
 CURRENT APPLICATION NUMBER: US/10/398,519  
 CURRENT FILING DATE: 2003-04-04  
 PRIOR APPLICATION NUMBER: 60/237,858  
 PRIOR FILING DATE: 2000-10-04  
 NUMBER OF SEQ ID NOS: 33  
 SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 14  
 LENGTH: 263  
 TYPE: PR  
 ORGANISM: Mus sp.  
 US-10-398-519-14

Query Match 78.4%; Score 1064.5; DB 12; Length 263;  
 Best Local Similarity 77.6%; Pred. No. 9,8e-69;  
 Matches 208; Conservative 16; Mismatches 17; Indels 27; Gaps 9;

QY 1 MADNFSIHDALSGSGNPNFQGWPGAWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 1 MADSEFLNDALHSGSGNPNFQGWPGAWGNOP-GAGGYPGAYPGAYPGQAPP 59  
 QY 60 GAYHG-----APGAYPGAPAPGYPPGPGGAYPPSSGQPSAPGAY----- 100  
 DB 60 GAYPGQAPPSAYPGPTAPGAYPGPTAPGAYPGQPA-PGAFPP-GQPGAGAYPGQSGGYP 116  
 QY 101 ATGPYCAPAGPLIVPYNLPLPGGVPRMLITLIGTVKPNANRIALDPRGNDVAFHF-PR 159  
 DB 117 AAGP-GVPAGPLIVPYDLPLPGGVPRMLITLIGTVKPNANRIALDPRGNDVAFHFPR 175  
 QY 160 FNNENRIVICNTKLNMMGREGROSVFPESGKPFKIQVLVEPDHFKVAVNDAAH-LQYN 218  
 DB 176 FNNENRIVICNTKODNNMGKEERQSAFPESGKPFKIQVLVEADHFKVAVNDAAHLLQYN 235  
 QY 219 HRVKKLNEISKLGISGDIDITLSASYTMI 246  
 DB 236 HRMKNLREISQIGISGDITLSANHAM 263

RESULT 19  
 US-10-133-234A-6  
 ; Sequence 6, Application US/10133234A  
 ; Publication No. US20040071684A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Panjwani et al.  
 ; TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the  
 ; FILE REFERENCE: 2002458-0006  
 ; CURRENT APPLICATION NUMBER: US/10/133,234A  
 ; CURRENT FILING DATE: 2002-04-26  
 ; NUMBER OF SEQ ID NOS: 28  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 245  
 ; TYPE: PR  
 ; ORGANISM: Hamster  
 ; US-10-133-234A-6

Query Match 77.3%; Score 1049.5; DB 12; Length 245;  
 Best Local Similarity 81.0%; Pred. No. 1,1e-67;  
 Matches 204; Conservative 12; Mismatches 23; Indels 13; Gaps 8;

QY 1 MAONFSIHDALSGSGNPNFQGWPGAWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 1 MADSEFLNDALHSGSGNPNFQGWPGAWGNOP-GAGGYPGAYPGAYPGQAPP 59  
 QY 60 GAYHG-APGAYPGAPAPGYPPGPGGAYPPSSGQPSAPGAY-ATGPYCAPAGPLIVPY 116  
 DB 60 GAYPGQAPPSAYPGPTAPGAYPGPTAPGAYPGQPA-PGAFPP-GQPGAGAYPGQSGGYP 116  
 QY 117 NLPLPGGVPRMLITLIGTVKPNANRIALDPRGNDVAFHF-PRFNNENRIVICNTKLD 175  
 DB 114 KLPLAGGVPRMLITLIGTVKPNANRIALDPRGNDVAFHFPRFNNENRIVICNTKLD 173  
 QY 176 NMMGREGROSVFPESGKPFKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNEISKLGISG 234  
 DB 174 NMMGREGROSAFPESGKPFKIQVLVEADHFKVAVNDAAHLLQYNHRMKNLREINQMSISG 233  
 QY 235 DIDITLSASYTMI 246  
 DB 234 DITLSASYTMI 245

RESULT 20  
 US-10-133-234A-4  
 ; Sequence 4, Application US/10133234A  
 ; Publication No. US20040071684A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Panjwani et al.  
 ; TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the  
 ; FILE REFERENCE: 2002458-0006  
 ; CURRENT APPLICATION NUMBER: US/10/133,234A  
 ; CURRENT FILING DATE: 2002-04-26  
 ; NUMBER OF SEQ ID NOS: 28  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 242  
 ; TYPE: PR  
 ; ORGANISM: Rabbit  
 ; US-10-133-234A-4

Query Match 75.5%; Score 1024; DB 12; Length 242;  
 Best Local Similarity 78.8%; Pred. No. 7,3e-66;  
 Matches 197; Conservative 15; Mismatches 26; Indels 12; Gaps 7;

QY 1 MADNFSIHDALSGSGNPNFQGWPGAWGNOPAGAGYPGASYPG-YPGQAPPGAYPGQAPP 59  
 DB 1 MADSEFLNDALHSGSGNPNFQGWPGAWGNOPAGGYPGAYPGAYPGQAPP 59  
 QY 60 GAYHGAPGAPAPGYPPGPGGAYPPSSGQPSAPGAY-ATGPYCAPAGPLIVPY 118  
 DB 60 GYPYG-PGAA-----GAYPGQPGGAYPPGQPGAGAYPGASYSASAGPLPVYDL 112  
 QY 119 PLPGGVPRMLITLIGTVKPNANRIALDPRGNDVAFHF-PRFNNENRIVICNTKLDNN 177  
 DB 113 PLPGGVPRMLITLIGTVKPNANRIALDPRGNDVAFHFPRFNNENRIVICNTKLDNN 172  
 QY 178 WGREERQSVFPESGKPFKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNEISKLGISGDI 236  
 DB 173 WGREERQSVFPESGKPFKIQVLVEPDHFKVAVNDAAHLLQYNHRMKNLREINQMSISGDI 232  
 QY 237 DITLSASYTMI 246  
 DB 233 QLTSAASHAMI 242

RESULT 21  
 US-09-728-479-6  
 ; Sequence 6, Application US/09728479  
 ; Patent No. US20020034726A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KANEKASAKI, SHIRO  
 ; APPLICANT: MATSUMOTO, RYOJI  
 ; APPLICANT: HIRASHIMA, MITSUOMI  
 ; TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR  
 ; FILE REFERENCE: 3914-2  
 ; CURRENT APPLICATION NUMBER: US/09/728,479  
 ; CURRENT FILING DATE: 2001-08-16  
 ; PRIOR APPLICATION NUMBER: PCT/JP99/02952  
 ; PRIOR FILING DATE: 1999-06-02  
 ; PRIOR APPLICATION NUMBER: JP 10/170698  
 ; PRIOR FILING DATE: 1998-06-02  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 149  
 ; TYPE: PR  
 ; ORGANISM: Homo sapiens  
 ; US-09-728-479-6

Query Match 55.4%; Score 752; DB 9; Length 149;  
 Best Local Similarity 98.6%; Pred. No. 1,5e-46;  
 Matches 146; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 10 ATPEYGAPEGELLVPEVNLPEGVVPRMLITLLGVKVEVIANIALDQFGNDVAFHH-PR 169

DQ 2 ATPEYGAPEGELLVPEVNLPEGVVPRMLITLLGVKVEVIANIALDQFGNDVAFHHNPR 61

QY 160 FNNENRRVIVCNTKLNMMGREEROSVPEFESGKPEKIQVLVEPHEFVAVNDAH-LQYN 218

DQ 62 FNNENRRVIVCNTKLNMMGREEROSVPEFESGKPEKIQVLVEPHEFVAVND AHLQYN 121

QY 219 HRVVKLNEISKLGISGIDILTSASYTMI 246

DQ 122 HRVVKLNEISKLGISGIDILTSASTMI 149

```

RESULT 22
US-09-877-790-1
; Sequence 1, Application US/09877790
; Publication No. US20030054982A1
; GENERAL INFORMATION:
; APPLICANT: Jarvis, Gary
; APPLICANT: John, Constance
; APPLICANT: Leffler, Hakon
; TITLE OF INVENTION: N-TERMINALLY TRUNCATED GALECTIN-3 FOR USE IN TREATING CANCER
; FILE REFERENCE: 3157.00004
; CURRENT APPLICATION NUMBER: US/09/877,790
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 143
; TYPE: PRT
; ORGANISM: homo sapien
; US-09-877-790-1

```

RESULT 23  
US-10-409-786-1  
Sequence 1, Application US/10/409786  
Publication No. US2004002855A1  
GENERAL INFORMATION:  
APPLICANT: John, Constance  
APPLICANT: Unger, Gretchen  
TITLE OR INVENTION: BIOLOGIC MODULATIONS WITH NANOPARTICLES  
FILE REFERENCE: 3157.00009  
CURRENT APPLICATION NUMBER: US/10/409,786  
CURRENT FILING DATE: 2003-04-08  
PRIOR APPLICATION NUMBER: 10/378,044  
PRIOR FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: 60/334,315  
PRIOR FILING DATE: 2002-07-08  
PRIOR APPLICATION NUMBER: 60/370,882  
PRIOR FILING DATE: 2002-04-08  
PRIOR APPLICATION NUMBER: 60/428,296  
PRIOR FILING DATE: 2002-11-22  
PRIOR APPLICATION NUMBER: 09/877,790  
PRIOR FILING DATE: 2001-06-08  
PRIOR APPLICATION NUMBER: WO 02/100343

```

; PRIOR FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 1
; LENGTH: 143
; TYPE: PR1
; ORGANISM: homo sapien
US-10-409-786-1

Query Match: 53.3% Score 723; DB 16; Length 143;
Best Local Similarity 98.6% Pred. No. 1.7e+4;
Matches 141; Conservative 0; Mismatches 0; Indels 2; Gaps 2

```

QY 100 GAAGAGLLIVYMLPLPGGVSPRMLTTIIGTAKPPANNRALDFOGQGNVAEHF - PRENN 164

Db 1 GAPAGGLIVYMLPLPGGVSPRMLTTIIGTAKPPANNRALDFOGQGNVAEHENPRENN 60

QY 165 RRIYVNTKLDNNMREEROSVFPPESGPKRIQVLVEPDHFKAVANDAH - LOYNHVVK 223

Db 61 RRIYVNTKLDNNMREEROSVFPPESGPKRIQVLVEPDHFKAVANDAHLOYNHVRK 120

QY 224 LNEISKLGISGDIDLTASASYMI 246

Db 121 LNEISKLGISGDIDLTASASYMI 143

```

RESULT 25
US-10-133-234A-5
; Sequence 5, Application US/10133234A
; Publication No. US20040071684A1
; GENERAL INFORMATION:
; APPLICANT: Panjwani et al
; TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the
; TITLE OF INVENTION: Re-epithelialization of wounds
; FILE REFERENCE: 2002458-0006
; CURRENT APPLICATION NUMBER: US/10/133,234A
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 262
; TYPE: prt
; ORGANISM: Chicken
; US-10-133-234A-5

```

Query Match	51.3%;	Score 695.5;	DB 12;	Length 262;
Best Local Similarity	58.8%;	Pred. No. 3e-42;		
Matches	150;	Conservative	23;	Mismatches 51; Indels 31; Gaps 14;
QY	7	LEHDLSSG-----GNPNPGCMPGAMGQDPAGAGYTPGASVYGYGGQAP--PGAVPGQAF	58	
Db	24	LHPQLSDLLPAHNPAPPPQGM-----NRPDPGAF--APYPGAYGADGAPYPG--A	74	
QY	59	PGAVHGAGAYGAPAPAPVYGPSPGSPGAYSSGGPSAPGAY--ATGPGY-APAGPLTV	114	
Db	75	PGPHHGPPGPPGPGP-PPYPGGP--PGYP--GGP--GGPYPGGPTAPYSEAPAPLKV	127	
QY	115	PYNLPFGGVPRMLITLLGTGPKRANLIALDFQGNVVAHFH--PRENNNRVIVCMTK	173	
Db	128	PYDLPPLPGLMERLLITTTGTVNSPNPNSLDFKKGQDIAFFNPRFKEDHKRVIVCNSM	187	
QY	174	LDNNMGRERQSV--PPFESGKPFKIQVLVERDPHFYAVVNDH--LQYNHRVKKLNEISKLG	231	
Db	188	FQNNMGKEERTAPRPPFEPGTFPFLQVLCEGDHFYAVVNDALHQLFNFEKKLNGITKLC	247	
QY	232	ISGDIDLTSASYTMI	246	
Db	248	IAGDILTSVLTSMI	262	

```

RESULT 26
US-10-633-035-8
; Sequence 8, Application US/10633035
; Publication No. US20040068104A1
; GENERAL INFORMATION:
; APPLICANT: Seisi Kato
; APPLICANT: Yamaguchi Kimura
; APPLICANT: Shingo Sekine
; APPLICANT: Kouju Kamata
; TITLE OF INVENTION: HUMAN GALECTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: GIN-6707CPUS
; CURRENT APPLICATION NUMBER: US/10/633,035
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/485,951
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 9-226468
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03670
; PRIOR FILING DATE: 1998-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-633-035-8

```

Query Match	22.04%	Score 299	DB 12	Length 353
Best Local Similarity	37.58%	Pred. No. 1.2e-13		
Matches	69	Conservative	28	Mismatches 47; Indels 40; Gaps 8
QY	63	HGAFG--AYPGADPGVYGGPBGPGGAYBSSGQPSAGAYATGPGYGAAGFLVYNYLF	119	
DB	199	HSTGQGMFSTEGIP-FVYVYPRP-----AYT-----LPFTYF	228	
QY	120	LPGGVVRKMLITLITGYKVNPNRIALDFORGNDAVAFHF-PRFNNRNREVIYCNTEKLDNNW	178	
DB	229	IPNGIYRSKSLMISGNVLPDATRHNIRINRCGGDLAFHLNPRFNEK--AVVRNTQINNWS	285	
QY	179	GREERQSV-FPPESGKPEKIQVLVEPDHFKNYAVNDALH-QYNRVRKELNELSTKIGISD	235	
DB	286	GOEERSTLGRMPDSRSGQSFYWIICGHCIFVYANGQEMCEYHRLKNLDINTLITLVAGD	345	
QY	236	IDLIT	239	
DB	346	IQILT	349	

```

RESULT 27
US-09-728-479-11
; Sequence 11, Application US/09728479
; Patent No. US20020034726A1
; GENERAL INFORMATION:
; APPLICANT: KANEGASAKI, SHIRO
; APPLICANT: MATSUMOTO, RYOJI
; APPLICANT: HIRASHIMA, MITSUMI
; TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR
; FILE REFERENCE: 3914-2
; CURRENT APPLICATION NUMBER: US/09/728,479
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: PCT/JP99/02952
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: JP 10/170698
; PRIOR FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO. 11
; LENGTH: 322
; TYPE: prt
; ORGANISM: Mus sp.
US-09-728-479-11

```

```

Query Match Similarity 21.7% Score 295; DB 9; Length 322;
Best Local Similarity 37.0% Pred. No. 2e-13;
Matches 68; Conservative 29; Mismatches 47; Indels 40; Gaps 8;

QY 63 HGAGF--AYPGAPAGVYPCGPPSGAGAYPSGCGPAGCATATGATGYPAGPLIVPYLPE 119
DB 168 HSTPPQMTSTPGLP-PVYVPTP-----AAT-----LPFYTP 197

QY 120 LGGVAVPMILITLGTAVPNARIALDFQGNDAVFHF--PRENNNRREVIYCNTEKLDNNW 178
DB 198 LINGLYPKSSTMLSGVLPDPAIRFHINLRGSDILFHILNPRENN--AVVENTQINNSW 254

QY 179 GEEESQSV--FPFESGKPPKIQVLVEPDHFKAIVANDAH-QYNHVVKLTNELSKGIISGD 235
DB 255 GGEESILIGRMFPEVFGQSFSVWIIIEGHQCFKVAVVGQHCVEYIHLKNLDINTLITLVEAGD 314

QY 236 IDLT 239
DB 315 IQLT 318

```

RESULT 28  
US-10-415-586-1  
; Sequence 1, Application US/10415586  
; Publication No. US20040053346A1  
; GENERAL INFORMATION:  
; APPLICANT: HIRASHIMA, Mitsuo  
; APPLICANT: YAMAUCHI, Akira

```

; APPLICANT: KAGESHITA, Toshiro
; APPLICANT: NAKAMURA, Takamori
; APPLICANT: NISHI, No. US20040053346A1cmu
; TITLE OF INVENTION: Predicting agent for a metastasis
; FILE REFERENCE: 2003-0572A/MMC/01332
; CURRENT APPLICATION NUMBER: US/10/415,586
; PRIOR FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: PCT/JP01/09561
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: JP 2000-335077
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-415-586-1
```

```

Query Match      21.3%; Score 289; DB 12; Length 355;
Best Local Similarity 36.2%; Pred. No. 6.1e-13;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;
```

```

QY PGAYGAP-----PGVYGPSSG-----PGAYSSGP-----SAPGAYAT 102
DB PRVTVQPAFTVPSPGVCPPRRGRGRQKPPGVWPNPAPITQTVIHTVQSAPGQMS 209
QY 103 GP-----YGAPAGPLIVPNLPLPGVVPRLITITGVKPNRRLDFOGRNDVAFH 156
DB 210 TPALPMYHPHAYWM--PFTITLIGLYPSKSIILSGTVLPSAQRHINLCSGNHIAFH 267
QY 157 F-PRFNNRNRVIVCNKTLNNWGERQ--SVPEESGKPKIQVLVEPDHFKVAVNDA 213
DB 268 LNPRFDEN--AVVANTQIDNSWGSSEERSLPRKMPFVFGQSFVWILCEAHCLKVAVDQ 324
QY 214 HL-QYNHRVKKLNEISKIGSDIDL 239
DB 325 HLFETYHRLRNLPITNRLLEVGGDIQLT 351
```

```

RESULT 29
US-10-633-035-2
; Sequence 2, Application US/10633035
; Publication No. US20040068104A1
; GENERAL INFORMATION:
; APPLICANT: Seiei Kato
; APPLICANT: Yamaguchi Kimura
; APPLICANT: Shingo Sekine
; APPLICANT: Kouju Kamata
; TITLE OF INVENTION: HUMAN GALECTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE
; FILE REFERENCE: GIN-6707CPUS
; CURRENT APPLICATION NUMBER: US/10/633,035
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/485,951
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 9-226468
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03670
; PRIOR FILING DATE: 1998-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 2
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-633-035-2
```

```

Query Match      21.3%; Score 289; DB 12; Length 355;
Best Local Similarity 36.2%; Pred. No. 6.1e-13;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;
QY 66 PGAYGAP-----PGVYGPSSG-----PGAYSSGP-----SAPGAYAT 102
```

```

DB 150 PRVTVQPAFTVPSPGVCPPRRGRGRQKPPGVWPNPAPITQTVIHTVQSAPGQMS 209
QY 103 GP-----YGAPAGPLIVPNLPLPGVVPRLITITGVKPNRRLDFOGRNDVAFH 156
DB 210 TPALPMYHPHAYWM--PFTITLIGLYPSKSIILSGTVLPSAQRHINLCSGNHIAFH 267
QY 157 F-PRFNNRNRVIVCNKTLNNWGERQ--SVPEESGKPKIQVLVEPDHFKVAVNDA 213
DB 268 LNPRFDEN--AVVANTQIDNSWGSSEERSLPRKMPFVFGQSFVWILCEAHCLKVAVDQ 324
QY 214 HL-QYNHRVKKLNEISKIGSDIDL 239
DB 325 HLFETYHRLRNLPITNRLLEVGGDIQLT 351
```

```

RESULT 30
US-10-633-035-6
; Sequence 6, Application US/10633035
; Publication No. US20040068104A1
; GENERAL INFORMATION:
; APPLICANT: Seiei Kato
; APPLICANT: Yamaguchi Kimura
; APPLICANT: Shingo Sekine
; APPLICANT: Kouju Kamata
; TITLE OF INVENTION: HUMAN GALECTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE
; FILE REFERENCE: GIN-6707CPUS
; CURRENT APPLICATION NUMBER: US/10/633,035
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/485,951
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 9-226468
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03670
; PRIOR FILING DATE: 1998-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 6
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-633-035-6
```

```

Query Match      21.3%; Score 289; DB 12; Length 355;
Best Local Similarity 36.2%; Pred. No. 6.1e-13;
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;
```

```

QY 66 PGAYGAP-----PGVYGPSSG-----PGAYSSGP-----SAPGAYAT 102
DB PRVTVQPAFTVPSPGVCPPRRGRGRQKPPGVWPNPAPITQTVIHTVQSAPGQMS 209
QY 103 GP-----YGAPAGPLIVPNLPLPGVVPRLITITGVKPNRRLDFOGRNDVAFH 156
DB 210 TPALPMYHPHAYWM--PFTITLIGLYPSKSIILSGTVLPSAQRHINLCSGNHIAFH 267
QY 157 F-PRFNNRNRVIVCNKTLNNWGERQ--SVPEESGKPKIQVLVEPDHFKVAVNDA 213
DB 268 LNPRFDEN--AVVANTQIDNSWGSSEERSLPRKMPFVFGQSFVWILCEAHCLKVAVDQ 324
QY 214 HL-QYNHRVKKLNEISKIGSDIDL 239
DB 325 HLFETYHRLRNLPITNRLLEVGGDIQLT 351
```

```

RESULT 31
US-10-376-133-18
; Sequence 18, Application US/10376133
; Publication No. US20030165965A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: LGALS AS MODIFIERS OF THE CHK PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX03-014C
```



LENGTH: 378  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-144-649A-439

Query Match  
Best Local Similarity 21.3%; Score 289; DB 14; Length 378;  
36.2%; Pred. No. 6.5e-13;  
Matches 75; Conservative 29; Mismatches 65; Indels 38; Gaps 9;

QY 66 PGAYPGAPDA-----PGVYGPSPSG-----PGAYPSSGCP-----SARGAYAT 102  
DB 173 PRIVVQAFSTVPSPGVCPPRRGRGRKPPGWPANPAPITQTVHTVQSAFGQMF 232  
QY 103 GP-----YGAPAGPLIVPYNLPPLGGVPRMLITLITGVKNANRIALDFORGNDVAFH 156  
DB 233 TPAPPPMYPPHAYPM--PFTITLIGLVPKSKILLSTGLVPSAQRFHINCSGNHIAFH 290  
QY 157 F-PFENNRRVIVCNTKLDNNNGREERQ--SVPFESGKPKIOVLVEPDHFKAIVADA 213  
DB 291 LNPRDEN--AVANNTQIDNSWGSSEERSLPRKMPFVGQSFVWLCEAHCLKVAVDGQ 347  
QY 214 HL-QYNHRVKKLINEISKLIGSDIDL 239  
DB 348 HLFYVYHRLNLPITNRLLEVGDIO 374

RESULT 35  
US-10-133-234A-3  
Sequence 3, Application US/10133234A  
Publication No. US20040071684A1  
GENERAL INFORMATION:  
APPLICANT: Panjwani et al.  
TITLE OF INVENTION: Use of Galectin-3 and Galectin-7 to Promote the  
FILE REFERENCE: 2002458-0006  
CURRENT APPLICATION NUMBER: US/10/133,234A  
CURRENT FILING DATE: 2002-04-26  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 3  
LENGTH: 139  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-10-133-234A-3

Query Match  
Best Local Similarity 21.2%; Score 288; DB 12; Length 139;  
45.1%; Pred. No. 2.7e-13;  
Matches 65; Conservative 25; Mismatches 38; Indels 16; Gaps 8;

QY 111 PLIYPVNIPLPGVPRMLITLITGVK--NANRIALDFORG--NDVAFHF--PREME-- 162  
DB 1 PGVYALND-----GLKPGKTLIVKGTIVAPKNAKRAVINDGSKSEENDLVHFNREMEAH 56  
QY 163 NNREVVICNTKL--DNNNGREERQSVPFESGKPKIOVLVEPDHFKAIVADA--LYNH 219  
DB 57 GDQNTVVCNSKENDNEMGTQREARAFPFQMGQPFELISVIEDKFKVKNVNDGHEFRPH 116  
QY 220 RVKTLNEISKLGISGDDILTSASY 243  
DB 117 RL-KLEAVQYILGKIDIKLTSIKF 139

RESULT 36  
US-09-728-479-2  
Sequence 2, Application US/09728479  
Patent No. US20020034726A1  
GENERAL INFORMATION:  
APPLICANT: KANEASAKI, SHIRO  
APPLICANT: MATSUMOTO, RYOJI  
APPLICANT: HIRASHIMA, MITSUOMI

TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR  
FILE REFERENCE: 3914-2  
CURRENT APPLICATION NUMBER: US/09/728,479  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: PCT/JP99/02952  
PRIOR FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: JP 10/170698  
PRIOR FILING DATE: 1998-06-02  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 2  
LENGTH: 323  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-728-479-2

Query Match  
Best Local Similarity 20.9%; Score 283; DB 9; Length 323;  
34.1%; Pred. No. 1.5e-12;  
Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASVYGPQAPPGAYPGQAP-----GAYHGAAPGAPGAP--PGVYGPSPGAP 89  
DB 138 GSVGLSTISFQPPGWPANPAPITQTVHTVQSAFGQMFPAIPPMYHP-----AYP 192  
QY 90 SSGQPSAPGAYATGPYGAAPGLIVPYNLPPLGGVPRMLITLITGVKNANRIALDFOR 149  
DB 193 -----MPFTITLIGLVPKSKILLSTGLVPSAQRFHINCS 226  
QY 150 GNDVAFHF--PFENNRRVIVCNTKLDNNNGREERQ--SVPFESGKPKIOVLVEPDHF 206  
DB 229 GNHIAFHLNPRDEN--AVANNTQIDNSWGSSEERSLPRKMPFVGQSFVWLCEAHCL 285  
QY 207 KVAIVDAHL-QYNHRVKKLINEISKLIGSDIDL 239  
DB 286 KVAIVDQHLFELYHRLNLPITNRLLEVGDIO 319

RESULT 37  
US-10-415-586-2  
Sequence 2, Application US/10415586  
Publication No. US2004005346A1  
GENERAL INFORMATION:  
APPLICANT: HIRASHIMA, Mitsuomi  
APPLICANT: YAMAUCHI, Akira  
APPLICANT: KAGESHITA, Toshiro  
APPLICANT: NAKAMURA, Takamori  
APPLICANT: NISHI, No. US2004005346A1omu  
TITLE OF INVENTION: Predicting agent for a metastasis  
FILE REFERENCE: 2003-0572A/MMC/01332  
CURRENT APPLICATION NUMBER: US/10/415,586  
CURRENT FILING DATE: 2003-09-05  
PRIOR APPLICATION NUMBER: PCT/JP01/09561  
PRIOR FILING DATE: 2001-10-31  
PRIOR APPLICATION NUMBER: JP 2000-335077  
PRIOR FILING DATE: 2000-11-01  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 323  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-415-586-2

Query Match  
Best Local Similarity 20.9%; Score 283; DB 12; Length 323;  
34.1%; Pred. No. 1.5e-12;  
Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASVYGPQAPPGAYPGQAP-----GAYHGAAPGAPGAP--PGVYGPSPGAP 89  
DB 138 GSVGLSTISFQPPGWPANPAPITQTVHTVQSAFGQMFPAIPPMYHP-----AYP 192  
QY 90 SSGQPSAPGAYATGPYGAAPGLIVPYNLPPLGGVPRMLITLITGVKNANRIALDFOR 149

Db 193 -----MPFTTITIGIYPSKSLILSGTVLPASQRFHINCS 228  
QY 150 GNDVAFHF-PRENENNRVIVCNTKLDNNWGREERQ--SVPFESGKPKIQVLVEPDHF 206  
Db 229 GNHIAFHNLPRDEN---AVVANTQIDNSWGSEERSLPRKMPFVAGQSFVWLCEAHCL 285  
QY 207 KVAVNDAAH-QYNHRYVKKLNEISKIGSDIDLT 239  
Db 286 KVAVDGQHLFEYHRLNLPFTINRLEVGDDIQLT 319

RESULT 38  
US-10-024-298A-141  
; Sequence 141, Application US/10024298A  
; Publication No. US20030143540A1  
; GENERAL INFORMATION:  
; APPLICANT: ASAH KASEI KASHIHI KAISHA  
; APPLICANT: Goichi HONDA  
; APPLICANT: Shuji MURAMATSU  
; APPLICANT: Yukiko NAGANO  
; TITLE OF INVENTION: NF-K B Activating Gene  
; FILE REFERENCE: 1254-0191P  
; CURRENT APPLICATION NUMBER: US/10/024,298A  
; PRIOR FILING DATE: 2003-04-08  
; PRIOR APPLICATION NUMBER: 60/314,385  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/278,641  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: 60/258,315  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: JP254018/2001  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: JP008912/2001  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: JP402288/2000  
; PRIOR FILING DATE: 2000-12-28  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 141  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-024-298A-141

Query Match 20.9%; Score 283; DB 14; Length 323;  
Best Local Similarity 34.1%; Pred. No. 1.5e-12;  
Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPYQGAPGAVPGAP-----GAYHAGPAGYGAAP--PGVYPPSGPGAYP 89  
Db 138 GSVQLSYISFQPPGVPWNPAPITQTVIHTVQSAFGQSFPAIPMMYHPH-----AYP 192  
QY 90 SSGQSPAPGAYATGYPAGAPGLIVPNLPLPGVVPFPMILITIGTVKNANRIALDFQR 149  
Db 193 -----MPFTTITIGIYPSKSLILSGTVLPASQRFHINCS 228  
QY 150 GNDVAFHF-PRENENNRVIVCNTKLDNNWGREERQ--SVPFESGKPKIQVLVEPDHF 206  
Db 229 GNHIAFHNLPRDEN---AVVANTQIDNSWGSEERSLPRKMPFVAGQSFVWLCEAHCL 285  
QY 207 KVAVNDAAH-QYNHRYVKKLNEISKIGSDIDLT 239  
Db 286 KVAVDGQHLFEYHRLNLPFTINRLEVGDDIQLT 319

RESULT 39  
US-10-042-211A-141  
; Sequence 141, Application US/10042211A  
; Publication No. US20030170719A1  
; GENERAL INFORMATION:  
; APPLICANT: MATSUDA, Akio et al.  
; TITLE OF INVENTION: NFKB Activating Gene

FILE REFERENCE: 1254-0192P  
; CURRENT APPLICATION NUMBER: US/10/042,211A  
; PRIOR FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-402288  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: JP 2001-088912  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: JP 2001-254018  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/258,315  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/278,640  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: US 60/314,385  
; PRIOR FILING DATE: 2001-08-24  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 141  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-042-211A-141

Query Match 20.9%; Score 283; DB 14; Length 323;  
Best Local Similarity 34.1%; Pred. No. 1.5e-12;  
Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPYQGAPGAVPGAP-----GAYHAGPAGYGAAP--PGVYPPSGPGAYP 89  
Db 138 GSVQLSYISFQPPGVPWNPAPITQTVIHTVQSAFGQSFPAIPMMYHPH-----AYP 192  
QY 90 SSGQSPAPGAYATGYPAGAPGLIVPNLPLPGVVPFPMILITIGTVKNANRIALDFQR 149  
Db 193 -----MPFTTITIGIYPSKSLILSGTVLPASQRFHINCS 228  
QY 150 GNDVAFHF-PRENENNRVIVCNTKLDNNWGREERQ--SVPFESGKPKIQVLVEPDHF 206  
Db 229 GNHIAFHNLPRDEN---AVVANTQIDNSWGSEERSLPRKMPFVAGQSFVWLCEAHCL 285  
QY 207 KVAVNDAAH-QYNHRYVKKLNEISKIGSDIDLT 239  
Db 286 KVAVDGQHLFEYHRLNLPFTINRLEVGDDIQLT 319

RESULT 40  
US-10-617-217A-141  
; Sequence 141, Application US/10617217A  
; Publication No. US20040081986A1  
; GENERAL INFORMATION:  
; APPLICANT: MATSUDA, Akio et al.  
; TITLE OF INVENTION: NF-KB ACTIVATING GENE  
; FILE REFERENCE: 1254-0229P  
; CURRENT APPLICATION NUMBER: US/10/617,217A  
; PRIOR FILING DATE: 2003-07-11  
; PRIOR APPLICATION NUMBER: JP 2000-402288  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: JP 2001-088912  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: JP 2001-254018  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/258,315  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/278,640  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: US 60/314,385  
; PRIOR FILING DATE: 2001-08-24  
; NUMBER OF SEQ ID NOS: 224  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 141  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-617-217A-141



Query Match 20.9%; Score 283; DB 16; Length 323;  
 Best Local Similarity 34.1%; Pred. No. 1.5e-12;  
 Matches 73; Conservative 27; Mismatches 70; Indels 44; Gaps 8;

QY 38 GASYPGYGQAPPGAYPGQAPP-----GAYHGAPGAYGAPA--PGVYPPPGPGGAYP 89  
 DB 138 GSVGLSTYISFQPPGWPANPAPITQVIHTVQSAFGQMFSTRPALPMMYFPH-----AYP 192  
 QY 90 SSGGPPSAPGAYATGYPGAPAPLIVPNLPLPGGVVPRMLTTITGYKPNANRIALDPQR 149  
 DB 193 -----MPTITTLGGLYPSKSTLLSGTVLPSAQRPHINICS 228  
 QY 150 GNDVAFHF-PRFNENNRVIVCNTKLDNNMGREERQ--SVPPESGKPKIQVLVEPDHF 206  
 DB 229 GNHIAFHLNRFDEN---AVARNQIDNSWGSEERSLPRKMPFVRGGSFSVWILCEAHCL 285  
 QY 207 KVAVND AHL-QYNHRVKKLNEISKLGISGDI DLT 239  
 DB 286 KVAVDGQHLEFYVYHRLNRLPTINRLVEVGDIQLT 319

## RESULT 41

US-10-138-588-76  
 ; Sequence 76, Application US/10138588  
 ; Publication No. US20040018594A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alsobrook et al.  
 ; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACID  
 ; TITLE OF INVENTION: ENCODING THE ANTIGENS, AND METHODS OF USE  
 ; FILE REFERENCE: 21402-347A  
 ; CURRENT APPLICATION NUMBER: US/10/138, 588  
 ; PRIOR FILING DATE: 2002-05-01  
 ; PRIOR APPLICATION NUMBER: 60/288,395  
 ; PRIOR FILING DATE: 2001-05-03  
 ; PRIOR APPLICATION NUMBER: 60/308,901  
 ; PRIOR FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: 60/313,388  
 ; PRIOR FILING DATE: 2001-08-17  
 ; PRIOR APPLICATION NUMBER: 60/324,757  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: 60/288,900  
 ; NUMBER OF SEQ ID NOS: 203  
 ; SEQ ID NO 76  
 ; LENGTH: 355  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-138-588-76

Query Match 20.0%; Score 271; DB 15; Length 355;  
 Best Local Similarity 34.8%; Pred. No. 1.2e-11;  
 Matches 72; Conservative 30; Mismatches 67; Indels 38; Gaps 9;

QY 66 PGAYGAPA-----PGVYPPGPGS-----PGAYPSGQP-----SAPGAYAT 102  
 DB 150 PRTPVQPAESTVPFQVCPFPBRGRQKRPSPVRANPAPITQVIHTVQSAFGQMS 209  
 QY 103 GP-----YGAPAEPLIVPNLPLPGGVVPRMLTTITGYKPNANRIALDPQRGNDVAFH 156  
 DB 210 TPAIPPMYHPHAPFM--PFTTTPGGLYPSKSTLLSGTVLPSAQRPHINICSASHAFH 267  
 QY 157 F-PRFNENNRVIVCNTKLDNNMGREERQ--SVPPESGKPKIQVLVEPDHFVAVNDA 213  
 DB 268 MNRPFDEN---AVARNQIDNSWGSEERSLPRKMPFVRGGSFSVWILCEAHCLKVAVDGQ 324  
 QY 214 HL-QYNHRVKKLNEISKLGISGDI DLT 239  
 DB 325 HVEYVYHRLNRLPTINRLVEVGDIQLT 351

## RESULT 42

US-10-633-035-7

Sequence 7, Application US/10633035  
 ; Publication No. US20040068104A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Selsi Kato  
 ; APPLICANT: Yamaguchi Kimura  
 ; APPLICANT: Shingo Sekine  
 ; APPLICANT: Kouju Kamata  
 ; TITLE OF INVENTION: HUMAN GALACTIC-9-LIKE PROTEINS AND CDNA ENCODING THESE  
 ; TITLE OF INVENTION: PROTEINS  
 ; FILE REFERENCE: GIN-6707CPIUS  
 ; CURRENT APPLICATION NUMBER: US/10/633, 035  
 ; PRIOR FILING DATE: 2003-08-04  
 ; PRIOR APPLICATION NUMBER: US/09/485,951  
 ; PRIOR FILING DATE: 2000-02-17  
 ; PRIOR APPLICATION NUMBER: 9-226468  
 ; PRIOR FILING DATE: 1997-08-22  
 ; PRIOR APPLICATION NUMBER: PCT/JP98/03670  
 ; PRIOR FILING DATE: 1998-08-19  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 7  
 ; LENGTH: 322  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-633-035-7

Query Match 19.7%; Score 267; DB 12; Length 322;  
 Best Local Similarity 33.2%; Pred. No. 2.1e-11;  
 Matches 71; Conservative 27; Mismatches 72; Indels 44; Gaps 8;

QY 38 GASYPGYGQAPPGAYPGQAPP-----GAYHGAPGAYGAPA--PGVYPPPGPGGAYP 89  
 DB 137 GSVGLSTYISFQPPGWPANPAPITQVIHTVQSAFGQMFSTRPALPMMYFPH-----AYP 191  
 QY 90 SSGGPPSAPGAYATGYPGAPAPLIVPNLPLPGGVVPRMLTTITGYKPNANRIALDPQR 149  
 DB 192 -----MPTITTLGGLYPSKSTLLSGTVLPSAQRPHINICS 227  
 QY 150 GNDVAFHF-PRFNENNRVIVCNTKLDNNMGREERQ--SVPPESGKPKIQVLVEPDHF 206  
 DB 228 GNHIAFHLNRFDEN---AVARNQIDNSWGSEERSLPRKMPFVRGGSFSVWILCEAHCL 284  
 QY 207 KVAVND AHL-QYNHRVKKLNEISKLGISGDI DLT 239  
 DB 285 KVAVDGQHLEFYVYHRLNRLPTINRLVEVGDIQLT 318

## RESULT 43

US-09-728-479-12  
 ; Sequence 12, Application US/09728479  
 ; Patent No. US20020034726A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KANEGASAKI, SHIRO  
 ; APPLICANT: MATSUMOTO, RYOJI  
 ; APPLICANT: HIRASHIMA, MITSUOMI  
 ; TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR  
 ; FILE REFERENCE: 3914-2  
 ; CURRENT APPLICATION NUMBER: US/09/728,479  
 ; PRIOR FILING DATE: 2001-08-16  
 ; PRIOR APPLICATION NUMBER: PCT/JP99/02952  
 ; PRIOR FILING DATE: 1999-06-02  
 ; PRIOR APPLICATION NUMBER: JP 10/170698  
 ; PRIOR FILING DATE: 1998-06-02  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 12  
 ; LENGTH: 323  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-728-479-12

Query Match 19.7%; Score 267; DB 9; Length 323;  
 Best Local Similarity 33.2%; Pred. No. 2.1e-11;

Matches 71; Conservative 27; Mismatches 72; Indels 44; Gaps 8;

QY 38 GASYPGYGAAPPAGYPCQAPP-----GAYHAGAGAYGAPAA--PGYPPGPPSGPAGY 89  
DB 138 GSVGLSYISFQPPGVPANPAPITQTIVHTVQSAGQWSTFPAIPPMYFHP-----AYP 192

QY 90 SSGQPSAAGAAATGTYGAPAGLIYPNYLPPLPGGVPPMLITITIGTYKPNRILADFOR 149  
DB 193 -----MPFTITIGSLYPSKSTILSGTLPSSAQRPHINLCS 228

QY 150 GNDVAFHP--PENNNRRVIVCNTKLDNNMGREERQ--SVPPESGKPEFKIQIVLVEPDHF 206  
DB 229 GNIHAFHILNRDEN---AVRNTOIDNSWSESESLPRKMPFVAGQSFSWILIGACIL 285

QY 207 KVAVNDAAH--QYHFRVKKLNEISKLGISGDIIDL 239  
DB 286 KVAVDGQHLFEYHYHRLNLPITNRLVGVGDIOL 319

RESULT 44  
US-09-263-689-4  
; Sequence 4, Application US/09263689  
; Patent No. US20020150970A1  
; GENERAL INFORMATION:  
; APPLICANT: NI, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/263,689  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/946,914  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 311 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-263-689-4

Query Match 19.5%; Score 264; DB 9; Length 311;  
Best Local Similarity 36.5%; Pred. No. 3.3e-11;  
Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 AGAAYGAPAA--PGYPPGPPSGGAPSSGQPSAPGAYATGTYGAPAGLIYPNYLPPLG 122  
DB 159 APQGFSTFPAIPPMYFHP-----AYP-----MPFTITIG 189

QY 123 GVPFMLITITIGTYKPNRILADFORGNDVAFHF--PENNNRRVIVCNTKLDNNMGRE 181  
DB 159 APQGFSTFPAIPPMYFHP-----AYP-----MPFTITIG 189

DB 190 GLYPSKSTILSGTLPSSAQRPHINLCSGNIHAFHILNRDEN---AVRNTOIDNSWSE 246

QY 182 ERQ--SVPPESGKPEFKIQIVLVEPDHFVAVNDAAH--QYHFRVKKLNEISKLGISGIDL 238  
DB 247 ERSILPRKMPFVAGQSFVWILCEAHCLKVAVDGQHLFEYHYHRLNLPITNRLVGVGDIOL 306

QY 239 T 239  
DB 307 T 307

RESULT 45  
US-10-415-586-3  
; Sequence 3, Application US/10415586  
; Publication No. US20040053346A1  
; GENERAL INFORMATION:  
; APPLICANT: HIRASHIMA, Mitsumori  
; APPLICANT: YAMAUCHI, Akira  
; APPLICANT: KAGESHITA, Toshio  
; APPLICANT: NAKAMURA, Takamori  
; APPLICANT: NISHI, No. US20040053346A1omv  
; TITLE OF INVENTION: Predicting agent for a metastasis  
; FILE REFERENCE: 2003-0572A/WMC/01332  
; CURRENT APPLICATION NUMBER: US/10/415,586  
; CURRENT FILING DATE: 2003-09-05  
; PRIOR APPLICATION NUMBER: PCT/JF01/09561  
; PRIOR FILING DATE: 2001-10-31  
; PRIOR APPLICATION NUMBER: JP 2000-335077  
; PRIOR FILING DATE: 2000-11-01  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 3  
; LENGTH: 311  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-415-586-3

Query Match 19.5%; Score 264; DB 12; Length 311;  
Best Local Similarity 36.5%; Pred. No. 3.3e-11;  
Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 AGAAYGAPAA--PGYPPGPPSGGAPSSGQPSAPGAYATGTYGAPAGLIYPNYLPPLG 122  
DB 159 APQGFSTFPAIPPMYFHP-----AYP-----MPFTITIG 189

QY 123 GVPFMLITITIGTYKPNRILADFORGNDVAFHF--PENNNRRVIVCNTKLDNNMGRE 181  
DB 190 GLYPSKSTILSGTLPSSAQRPHINLCSGNIHAFHILNRDEN---AVRNTOIDNSWSE 246

QY 182 ERQ--SVPPESGKPEFKIQIVLVEPDHFVAVNDAAH--QYHFRVKKLNEISKLGISGIDL 238  
DB 247 ERSILPRKMPFVAGQSFVWILCEAHCLKVAVDGQHLFEYHYHRLNLPITNRLVGVGDIOL 306

QY 239 T 239  
DB 307 T 307

RESULT 46  
US-10-235-674-4  
; Sequence 4, Application US/10235674  
; Publication No. US20030040081A1  
; GENERAL INFORMATION:  
; APPLICANT: NI, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 8 and 10SV Polynucleotides  
; FILE REFERENCE: 1488.0560004  
; CURRENT APPLICATION NUMBER: US/10/235,674  
; CURRENT FILING DATE: 2002-09-06  
; PRIOR APPLICATION NUMBER: US 09/656,450  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: US 09/263,689

; PRIOR FILING DATE: 1999-03-05  
 ; PRIOR APPLICATION NUMBER: US 08/946,914  
 ; PRIOR FILING DATE: 1997-10-09  
 ; PRIOR APPLICATION NUMBER: US 60/028,093  
 ; PRIOR FILING DATE: 1996-10-09  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 311  
 ; TYPE: PR1  
 ; ORGANISM: Homo sapiens  
 ; US-10-235-674-4

Query Match 19.5%; Score 264; DB 14; Length 311;  
 Best Local Similarity 36.5%; Pred. No. 3.3e-11;  
 Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 APGAYGAP--FGYPPPPGPGGAYPSSGGPSAPGAYATGPGAPAGPLIVPYNLPJG 122  
 Db 159 APGWFTPAIPPMYPPH-----AYP-----MPFITILG 189  
 QY 123 GVPRMLTITIGTYKPNRRLALDQGNDAVPHF--PRENNRRVIVCNTRLNNMRE 181  
 Db 190 GLYSKSLISGTVLPQRPINICSNHIAFHINPRDEN--AVRNTQIDNSWGSE 246  
 QY 182 ERQ--SVPPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKKNEISKLGSDIDL 238  
 Db 247 ERSLPRKMPFVRGGSFSVWLICFAHCLKVAVDGQHLFEYHRLRLPTINLLEVAGDIQL 306  
 QY 239 T 239  
 Db 307 T 307

RESULT 47  
 US-09-728-479-8  
 ; Sequence 8, Application US/09728479  
 ; Patent No. US20020034726A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KANEGASAKI, SHIRO  
 ; APPLICANT: MATSUMOTO, RYOJI  
 ; APPLICANT: HIRASHIMA, MITSUOMI  
 ; TITLE OF INVENTION: EOSINOPHIL CHEMOTACTIC FACTOR  
 ; FILE REFERENCE: 3914-2  
 ; CURRENT APPLICATION NUMBER: US/09/728,479  
 ; PRIOR FILING DATE: 2001-08-16  
 ; PRIOR APPLICATION NUMBER: PCT/JP99/02952  
 ; PRIOR FILING DATE: 1999-06-02  
 ; PRIOR APPLICATION NUMBER: JP 10/170698  
 ; PRIOR FILING DATE: 1998-06-02  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 8  
 ; LENGTH: 145  
 ; TYPE: PR1  
 ; ORGANISM: Rattus sp.  
 ; US-09-728-479-8

Query Match 18.5%; Score 250.5; DB 9; Length 145;  
 Best Local Similarity 40.4%; Pred. No. 1.4e-10;

Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;  
 QY 108 PAGPLIVPYNLPJGQVPRMLTITIGTYKPNRRLALDQGNDAVPHF--PRENNRR 166  
 Db 9 PYPLAVPFPTISNGLYPSKSIYISGVVLSDAKRFQINLRCGGDIAPHILNPRDEN-- 65  
 QY 167 VIVCNTRLNNMNGEERQ--SVPPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKK 223  
 Db 66 AVVENTQINNSWGEERSLPQSMFPRGQPSVWLCEGHCFKVAVDGQHICEYSHRLMN 125  
 QY 224 INEISKLGSDIDL 239  
 Db 126 LPDINTLEVAGDIQL 141

# RESULT 48

US-09-894-526-5  
 ; Sequence 5, Application US/09894526  
 ; Patent No. US20020127689A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.  
 ; Goli, Surya K.  
 ; Bandman, Olga  
 ; Hawkins, Phillip R.  
 ; Petithory, Joanne R.

TITLE OF INVENTION: NOVEL HUMAN GALECTINS  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Discrete  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/894,526  
 FILING DATE: 27-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/788,584  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PR-0192 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 145 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: GenBank  
 CLONE: 727176

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 ; US-09-894-526-5

Query Match 18.5%; Score 250.5; DB 9; Length 145;  
 Best Local Similarity 40.4%; Pred. No. 1.4e-10;  
 Matches 55; Conservative 26; Mismatches 48; Indels 7; Gaps 4;

QY 108 PAGPLIVPYNLPJGQVPRMLTITIGTYKPNRRLALDQGNDAVPHF--PRENNRR 166  
 Db 9 PYPLAVPFPTISNGLYPSKSIYISGVVLSDAKRFQINLRCGGDIAPHILNPRDEN-- 65  
 QY 167 VIVCNTRLNNMNGEERQ--SVPPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKK 223  
 Db 66 AVVENTQINNSWGEERSLPQSMFPRGQPSVWLCEGHCFKVAVDGQHICEYSHRLMN 125  
 QY 224 INEISKLGSDIDL 239  
 Db 126 LPDINTLEVAGDIQL 141

# RESULT 49

US-09-263-689-12  
 ; Sequence 12, Application US/09263689  
 ; Patent No. US20020150970A1

```

1 GENERAL INFORMATION:
2 APPLICANT: Ni, Jian
3 APPLICANT: Gentz, Reiner L.
4 APPLICANT: Ruben, Steven M.
5 TITLE OF INVENTION: Galeclin 8, 9, 10 and 10SV
6 NUMBER OF SEQUENCES: 60
7 CORRESPONDENCE ADDRESS:
8 ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
9 STREET: 1100 New York Ave., Suite 600
10 CITY: Washington
11 STATE: D.C.
12 COUNTRY: USA
13 ZIP: 20005-3914
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.30
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/09/263,689
21 FILING DATE:
22 CLASSIFICATION:
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: 08/946,914
25 FILING DATE:
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Steffe, Eric K.
28 REGISTRATION NUMBER: 36,688
29 REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 202-371-2600
32 TELEFAX: 202-371-2540
33 INFORMATION FOR SEQ ID NO: 12:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 145 amino acids
36 TYPE: amino acid
37 STRANDNESS: not relevant
38 TOPOLOGY: linear
39 MOLECULE TYPE: cdna
40 IS-09-263-689-12

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 14:36:22 ; Search time 22 Seconds

(without alignments)  
577.272 Million cell updates/sec

Title: US-09-297-040-4

Perfect score: 1357

Sequence: 1 MADNPSLHDAISGSGNPPO.....ISKLGISGIDIDITSASYTMI 246

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :  
1: Issued Patents AA.\*  
2: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/6C.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1315	96.9	250	3	US-08-946-914-10
2	1315	96.9	250	4	US-09-656-450-10
3	1305	96.2	250	1	US-08-562-311-2
4	1079	79.5	262	3	US-08-946-914-14
5	1079	79.5	262	4	US-09-656-450-14
6	1078	79.4	264	1	US-08-562-311-4
7	513	37.8	135	2	US-08-647-960-5
8	277.5	20.4	322	4	US-09-559-023-2
9	264	19.5	311	3	US-08-946-914-4
10	264	19.5	311	4	US-09-656-450-4
11	250.5	18.5	145	2	US-08-788-584-5
12	250.5	18.5	145	3	US-08-946-914-12
13	250.5	18.5	145	4	US-09-656-450-12
14	250.5	18.5	145	5	US-09-557-170A-3
15	249	18.3	149	2	US-08-788-584-3
16	246.5	18.2	168	4	US-09-401-064-199
17	243.5	17.9	145	2	US-08-788-584-1
18	243.5	17.9	301	4	US-09-559-023-4
19	236.5	17.4	324	3	US-08-946-914-11
20	236.5	17.4	324	4	US-09-656-450-11
21	223.5	16.5	200	3	US-08-946-914-8
22	223.5	16.5	200	4	US-09-656-450-8
23	223.5	16.5	316	3	US-09-131-648-5
24	223.5	16.5	317	3	US-08-946-914-6
25	223.5	16.5	317	4	US-09-656-450-6
26	218.5	16.1	316	2	US-08-728-521-3
27	218.5	16.1	316	2	US-08-647-960-2

28	218.5	16.1	316	3	US-08-946-914-15	Sequence 15, Appl
29	218.5	16.1	316	3	US-08-946-914-17	Sequence 17, Appl
30	218.5	16.1	316	3	US-09-212-146-3	Sequence 3, Appl
31	218.5	16.1	316	4	US-09-656-450-15	Sequence 15, Appl
32	218.5	16.1	316	4	US-09-656-450-17	Sequence 17, Appl
33	215.5	15.9	323	3	US-08-469-667-16	Sequence 16, Appl
34	215.5	15.9	323	3	US-08-946-914-2	Sequence 2, Appl
35	215.5	15.9	323	4	US-09-224-110-16	Sequence 16, Appl
36	215.5	15.9	323	4	US-09-656-450-2	Sequence 2, Appl
37	215.5	15.9	323	5	PCT-US95-07289-16	Sequence 16, Appl
38	211	15.5	136	3	US-08-946-914-13	Sequence 13, Appl
39	211	15.5	136	4	US-09-154-750A-79	Sequence 79, Appl
40	211	15.5	136	4	US-09-656-450-13	Sequence 13, Appl
41	207	15.3	318	4	US-09-557-170A-4	Sequence 4, Appl
42	200.5	14.8	504	3	US-09-219-849-3	Sequence 3, Appl
43	200.5	14.8	561	1	US-08-642-255-52	Sequence 52, Appl
44	199.5	14.7	144	1	US-08-642-255-49	Sequence 49, Appl
45	199.5	14.7	720	3	US-09-219-849-4	Sequence 4, Appl
46	199.5	14.7	777	1	US-08-642-255-53	Sequence 53, Appl
47	197.5	14.6	234	1	US-08-642-255-51	Sequence 51, Appl
48	196.5	14.5	466	3	US-08-526-136-13	Sequence 13, Appl
49	195.5	14.4	417	1	US-08-175-155-69	Sequence 69, Appl
50	195.5	14.4	417	1	US-08-477-509B-104	Sequence 104, Appl
51	195.5	14.4	417	1	US-08-642-255-102	Sequence 102, Appl
52	195.5	14.4	417	2	US-08-707-237A-76	Sequence 76, Appl
53	195.5	14.4	417	3	US-08-482-085B-104	Sequence 104, Appl
54	195.5	14.4	417	4	US-09-444-791A-104	Sequence 104, Appl
55	195.5	14.4	829	1	US-08-642-255-132	Sequence 132, Appl
56	195.5	14.4	829	1	US-08-397-633A-53	Sequence 53, Appl
57	195.5	14.4	837	1	US-08-175-155-68	Sequence 68, Appl
58	195.5	14.4	837	1	US-08-477-509B-103	Sequence 103, Appl
59	195.5	14.4	837	1	US-08-642-255-101	Sequence 101, Appl
60	195.5	14.4	837	2	US-08-707-237A-75	Sequence 75, Appl
61	195.5	14.4	837	3	US-08-482-085B-103	Sequence 103, Appl
62	195.5	14.4	837	4	US-09-444-791A-103	Sequence 103, Appl
63	195.5	14.4	897	1	US-08-397-633A-50	Sequence 50, Appl
64	192	14.1	166	4	US-09-841-334A-21	Sequence 21, Appl
65	191	14.1	160	3	US-08-542-051-18	Sequence 18, Appl

## ALIGNMENTS

RESULT 1  
US-08-946-914-10  
Sequence 10, Application US/08946914  
Patent No. 6027916  
GENERAL INFORMATION:  
APPLICANT: NI, Jian  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Galectin B, 9, 10 and 10SV  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
STREET: 1100 New York Ave., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946,914  
FILING DATE: Herewith  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,093  
FILING DATE: 09-OCT-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Steffe, Eric K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 250 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-946-914-10

Query Match 96.9%; Score 1315; DB 3; Length 250;  
Best Local Similarity 98.4%; Pred. No. 4,6e-114;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGWPGAWGNQPAAGAGYPPGASYPG-YPGQAPPGAYPGQAPP 59  
DB 1 MADNFSLHDALSGSGNPNPQGWPGAWGNQPAAGAGYPPGASYPGAYPGQAPPAYPGQAPP 60  
QY 60 GAYHAPGAYPGAPAPGAYPPGSPGAYPPSSGQSPAPGAY-ATGPYCAPPLIIVPYNL 118  
DB 61 GAYHAPGAYPGAPAPGAYPPGSPGAYPPSSGQSPAPGAYATGPYCAPPLIIVPYNL 120  
QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVFAHF-PFNENNRVIVCNKLDNN 177  
DB 121 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVFAHFNNRNNRIVCNKLDNN 180  
QY 178 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNISKLGISGDI 236  
DB 181 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAAHLLQYNHRVKKLNISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 2  
US-09-656-450-10  
Sequence 10, Application US/09656450  
Patent No. 6468768  
GENERAL INFORMATION:  
APPLICANT: NI, Jian  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides  
FILE REFERENCE: 1488.0560003  
CURRENT APPLICATION NUMBER: US/09/656,450  
CURRENT FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: US 09/263,689  
PRIOR FILING DATE: 1999-03-05  
PRIOR APPLICATION NUMBER: US 08/946,914  
PRIOR FILING DATE: 1997-10-09  
PRIOR APPLICATION NUMBER: US 60/028,093  
PRIOR FILING DATE: 1996-10-09  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 10  
LENGTH: 250  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-656-450-10

Query Match 96.9%; Score 1315; DB 4; Length 250;  
Best Local Similarity 98.4%; Pred. No. 4,6e-114;  
Matches 246; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGWPGAWGNQPAAGAGYPPGASYPG-YPGQAPPGAYPGQAPP 59  
DB 1 MADNFSLHDALSGSGNPNPQGWPGAWGNQPAAGAGYPPGASYPGAYPGQAPPAYPGQAPP 60

QY 60 GAYHAPGAYPGAPAPGAYPPGSPGAYPPSSGQSPAPGAY-ATGPYCAPPLIIVPYNL 118  
DB 61 GAYHAPGAYPGAPAPGAYPPGSPGAYPPSSGQSPAPGAYATGPYCAPPLIIVPYNL 120  
QY 119 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVFAHF-PFNENNRVIVCNKLDNN 177  
DB 121 PLPGGVPRMLITLIGTVKPNANRIALDFQGNDAVFAHFNNRNNRIVCNKLDNN 180  
QY 178 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAAH-LQYNHRVKKLNISKLGISGDI 236  
DB 181 WGERERQSVFPFESGKPKIQVLVEPDHFKVAVNDAAHLLQYNHRVKKLNISKLGISGDI 240  
QY 237 DLTSASYTMI 246  
DB 241 DLTSASYTMI 250

RESULT 3  
US-08-562-311-2  
Sequence 2, Application US/08562311  
Patent No. 5801002  
GENERAL INFORMATION:  
APPLICANT: RAZ, AVRAHAM  
TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY  
TITLE OF INVENTION: OF METASTASIS IN A CELL SAMPLE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dykema Goesselt  
STREET: STE 505 N. Woodward  
CITY: Bloomfield Hills  
STATE: MI  
COUNTRY: U.S.  
ZIP: 48304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/562,311  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,225  
FILING DATE:  
APPLICATION NUMBER: US 07/681,242  
FILING DATE: 04-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/294,249  
FILING DATE: 01-JUN-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: KELLY, ROBERT L.  
REGISTRATION NUMBER: 31,843  
REFERENCE/DOCKET NUMBER: 61,686-  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 810-540-0849  
TELEFAX: 810-540-0763  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 250 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-562-311-2

Query Match 96.2%; Score 1305; DB 1; Length 250;  
Best Local Similarity 97.6%; Pred. No. 3,9e-113;  
Matches 244; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 MADNFSLHDALSGSGNPNPQGWPGAWGNQPAAGAGYPPGASYPG-YPGQAPPGAYPGQAPP 59  
DB 1 MADNFSLHDALSGSGNPNPQGWPGAWGNQPAAGAGYPPGASYPGAYPGQAPPAYPGQAPP 60







SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 322  
 TYPE: PRT  
 ORGANISM: Rat  
 US-09-559-023-2

Query Match 20.4%; Score 277.5; DB 4; Length 322;  
 Best Local Similarity 32.9%; Pred. No. 6.9e-18;  
 Matches 82; Conservative 29; Mismatches 67; Indels 71; Gaps 12;

QY 5 FSLHDLGSG-----NPNQCPGAMGNQPAAGGYPGASVPGYCGAPGAYPQQA 57  
 DB 127 YHLVDTTISVSGCLHLSPINPQTQGF-----QPA-----HQA-----VAQT 162  
 QY 58 PPGAHGAPG---AVGAPAPGVYCPSPGAPSSGQPSAPGAYATGPAPAPGLIIV 114  
 DB 163 IITHVHSIFGQMLSTFGIIP-PMAYITP-----AYT-----I 192  
 QY 115 PYNLPDPGVVPRMLITLITGVKPNANRIALDFQGNDAFHF--PRENNRRVIVCNTK 173  
 DB 193 PFETSIPIGFGYPSKISINISGVLPDAKRFHINLRGGDIAPHINPREVE--KVVVNTQ 249  
 QY 174 LDNNWGREERQ--SVFPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKLNELISKL 230  
 DB 250 INNSWGEERSLPRMPFRNGQSPFSVWILCEHGFKVAVDGQHICEVYHRLKNLPDINTL 309  
 QY 231 GISGIDLT 239  
 DB 310 EVAGDIQLT 318

# RESULT 9

US-08-946-914-4  
 Sequence 4, Application US/08946914  
 Patent No. 6027916

GENERAL INFORMATION:  
 APPLICANT: NI, Jlan  
 APPLICANT: Gentz, Reiner L.  
 APPLICANT: Ruben, Steven M.  
 TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
 NUMBER OF SEQUENCES: 60  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Steirne, Kessler, Goldstein, & Fox P.L.L.C.  
 STREET: 1100 New York Ave., Suite 600  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3934

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/946,914  
 FILING DATE: Herewalt  
 CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/028,093  
 FILING DATE: 09-OCT-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Steffe, Eric K.  
 REGISTRATION NUMBER: 36,688  
 REFERENCE/DOCKET NUMBER: 1488, 0560001/EKS/SGM  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-371-2600  
 TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 311 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 US-08-946-914-4

Query Match 19.5%; Score 264; DB 3; Length 311;  
 Best Local Similarity 36.5%; Pred. No. 1.2e-16;  
 Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 APGAYPGAPA--PGVYGPSPGAPSSGQPSAPGAYATGPAPAPGLIIVPNLP 122  
 DB 159 APGGMSTPALPPMMYHP-----AYP-----MPTTTTLG 189  
 QY 123 GVVPRLITLITGVKPNANRIALDFQGNDAFHF--PRENNRRVIVCNTKLDNNWGRE 181  
 DB 190 GLYPSKSLISGLTLPDAQRFHINLCSGNHIAFLINRPDEN--AVVRNTQIDNSWGE 246  
 QY 182 ERQ--SVFPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKLNELISKLSG 238  
 DB 247 ERSIPRRKMPFVKGQSFVWILCEAHCLKVAVDGQHLFEYHRLNLPITNRLVGGDIQ 306  
 QY 239 T 239  
 DB 307 T 307

# RESULT 10

US-09-656-450-4  
 Sequence 4, Application US/09656450  
 Patent No. 6468768

GENERAL INFORMATION:  
 APPLICANT: NI, Jlan  
 APPLICANT: Gentz, Reiner L.  
 APPLICANT: Ruben, Steven M.  
 TITLE OF INVENTION: Galectin 9 and 10SV polynucleotides  
 FILE REFERENCE: 1488, 0560003  
 CURRENT APPLICATION NUMBER: US/09/656,450  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: US 09/263,689  
 PRIOR FILING DATE: 1999-03-05  
 PRIOR APPLICATION NUMBER: US 08/946,914  
 PRIOR FILING DATE: 1997-10-09  
 PRIOR APPLICATION NUMBER: US 60/028,093  
 PRIOR FILING DATE: 1996-10-09  
 NUMBER OF SEQ ID NOS: 60  
 SOFTWARE: Patentin version 3.0  
 SEQ ID NO 4  
 LENGTH: 311  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-656-450-4

Query Match 19.5%; Score 264; DB 4; Length 311;  
 Best Local Similarity 36.5%; Pred. No. 1.2e-16;  
 Matches 66; Conservative 25; Mismatches 52; Indels 38; Gaps 7;

QY 65 APGAYPGAPA--PGVYGPSPGAPSSGQPSAPGAYATGPAPAPGLIIVPNLP 122  
 DB 159 APGGMSTPALPPMMYHP-----AYP-----MPTTTTLG 189  
 QY 123 GVVPRLITLITGVKPNANRIALDFQGNDAFHF--PRENNRRVIVCNTKLDNNWGRE 181  
 DB 190 GLYPSKSLISGLTLPDAQRFHINLCSGNHIAFLINRPDEN--AVVRNTQIDNSWGE 246  
 QY 182 ERQ--SVFPESGKPKIQVLVEPDHFKVAVNDAHL-QYNHRVKLNELISKLSG 238  
 DB 247 ERSIPRRKMPFVKGQSFVWILCEAHCLKVAVDGQHLFEYHRLNLPITNRLVGGDIQ 306  
 QY 239 T 239  
 DB 307 T 307

RESULT 11  
 US-08-788-584-5





;; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE

;; FILE REFERENCE: 210121.471C2

;; CURRENT APPLICATION NUMBER: US/09/401.064

;; CURRENT FILING DATE: 1999-09-22

;; NUMBER OF SEQ ID NOS: 371

;; SOFTWARE: FastSeq for Windows Version 3.0

;; SEQ ID NO: 199

;; LENGTH: 168

;; TYPE: PRT

;; ORGANISM: Homo sapien

US-09-401-064-199

Query Match 18.2%; Score 246.5; DB 4; Length 168;  
Best Local Similarity 37.3%; Pred. No. 2.2e-15;

Matches 59; Conservative 33; Mismatches 51; Indels 15; Gaps 8;

QY 92 GQPSAPGAVATGPYAGAPLIVPNLPLPGGVPRMLITLITGV-KENANRIALDFQ-- 148

DB 5 GEMATSGQA--PYLSPA---VPFSGTIQGGIQQDGLQITVNGVITLSSGTRFAVNFQTG 58

QY 149 -RGNDVAFHF-PRFNNRNRVIVCNTKLDNMWGEREEROSVFPFESGKPFKIQVLEPDHF 206

DB 59 FSGNDIAHFHFPRFDDG--YVVCNTRQNGSWGPEERKTHMPFQKMPFDLCFIVQSSDF 116

QY 207 KVAVND-AHLOYNHRVKKLNEISKLGISGDIIDLTASV 243

DB 117 KVMVNGILFVQYFHFV-PPHRTVITSVNGSVOLSTISF 153

RESULT 17

US-08-788-584-1

Sequence 1, Application US/08788584

Patent No. 5837493

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Goli, Surya K.

APPLICANT: Bandman, Olga

APPLICANT: Hawkins, Phillip R.

APPLICANT: Peitchory, Joanne R.

TITLE OF INVENTION: NOVEL HUMAN GALECTINS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/788,584

FILING DATE: Filed Herewith

CLASSIFICATION: 436

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0192 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-788-584-1

Query Match 17.9%; Score 243.5; DB 2; Length 145;

Best Local Similarity 40.0%; Pred. No. 3.5e-15;

Matches 56; Conservative 23; Mismatches 50; Indels 11; Gaps 5;

QY 104 PYGAPAGPLIVPNLPLPGGVPRMLITLITGV-KENANRIALDFQ--RGNDVAFHF-PRNE 162

DB 9 PYLSPX---VPFSGTIQGGIYPSKSTILSGTVLPSAQRFHINLCGSHIGFHLNPRDE 64

QY 163 NNRRVIVCNTKLDNMWGEREERQ--SVFPFESGKPFKIQVLEPDHFVAVNDALH-QYNH 219

DB 65 N---AVVANNQIDNXXWGSERSLPRKMPFVRGQSPFVILCEAHCLKVAVDGQHLEFYH 121

QY 220 RVKLTNEISKLGISGDIIDLT 239

DB 122 RLRLPLTIRLEVGDIQLT 141

RESULT 18

US-09-559-023-4

Sequence 4, Application US/09559023

Patent No. 6551796

GENERAL INFORMATION:

APPLICANT: Abramson, Ruth

APPLICANT: Leal-Pinto, Edgar

APPLICANT: Lipkowitz, Michael

TITLE OF INVENTION: NUCLEIC ACID ENCODING URATE TRANSPORTER

FILE REFERENCE: 070165.0574

CURRENT APPLICATION NUMBER: US/09/559,023

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: US 09/221,898

PRIOR FILING DATE: 1998-12-28

PRIOR APPLICATION NUMBER: US 60/099,752

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: US 60/070,215

PRIOR FILING DATE: 1997-12-31

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 301

TYPE: PRT

ORGANISM: Human

US-09-559-023-4

Query Match 17.9%; Score 243.5; DB 4; Length 301;

Best Local Similarity 38.4%; Pred. No. 8.9e-15;

Matches 56; Conservative 30; Mismatches 47; Indels 13; Gaps 7;

QY 104 PYGAPAGPLIVPNLPLPGGVPRMLITLITGV-KENANRIALDFQ--RGNDVAFHF-P 158

DB 9 PYLSPA---VPFSGTIQGGIQQDGLQITVNGVITLSSGTRFAVNFQTG 64

QY 159 RFNNRNRVIVCNTKLDNMWGEREEROSVFPFESGKPFKIQVLEPDHFVAVND-AHLOY 217

DB 65 RREDG--YVVCNTRQNGSWGPEERKTHMPFQKMPFDLCFIVQSSDFVAVNGLIFVQY 122

QY 218 NRVYKLTNEISKLGISGDIIDLTASV 243

DB 123 FHRV-PPHRTVITSVNGSVOLSTISF 147

RESULT 19

US-08-946-914-11

Sequence 11, Application US/08946914

Patent No. 6027916

GENERAL INFORMATION:

APPLICANT: Ni, Jian

APPLICANT: Gentz, Reiner L.

APPLICANT: Ruben, Steven M.

TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV

NUMBER OF SEQUENCES: 60

```

CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-11

Query Match      17.4%; Score 236.5; DB 3; Length 324;
Best Local Similarity 35.2%; Pred. No. 4.3e-14;
Matches 68; Conservative 25; Mismatches 63; Indels 37; Gaps 10;

QY 71 GAPPAGVYPPSPGPGAYSSGQPSAPGAVATGPGAP-----AGPLI-----VPYNL 118
DB 152 GPPASQYPTMTIP-AVPSAG-----YNPPQMSLPVMAQPIFNFPVYVG 198
QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQRGN--DVAHF--PRENNRRVIVCNTKLD 175
DB 199 TLGGGLTARRITIIKGYVLPTAKNLIINFKVGSTGDIAPFHMNPRIGD---CVRANSYMN 254
QY 176 NNMGRERQSVF-PFESGKPEKIVLVEPDHFKAIVNDAH-QYNHRVKLINEISKLGIS 233
DB 255 GSWSEBKRKIPYNPFAGQGFDDLSIRCGTDRFKVFANGQHLDFSHRFQAFQVDMLEIK 314
QY 234 GDIDLTSASYTMI 246
DB 315 GDITL---SYVQI 324

```

## RESULT 20

```

US-09-656-450-11
Sequence 11, Application US/09656450
Patent No. 6468768
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
FILE REFERENCE: 1488.0560003
CURRENT APPLICATION NUMBER: US/09/656,450
CURRENT FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/263,689
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: US 08/946,914
PRIOR FILING DATE: 1997-10-09
PRIOR APPLICATION NUMBER: US 60/028,093

```

```

PRIOR FILING DATE: 1996-10-09
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Patent version 3.0
SEQ ID NO 11
LENGTH: 324
TYPE: PRT
ORGANISM: Rat
US-09-656-450-11

Query Match      17.4%; Score 236.5; DB 4; Length 324;
Best Local Similarity 35.2%; Pred. No. 4.3e-14;
Matches 68; Conservative 25; Mismatches 63; Indels 37; Gaps 10;

QY 71 GAPPAGVYPPSPGPGAYSSGQPSAPGAVATGPGAP-----AGPLI-----VPYNL 118
DB 152 GPPASQYPTMTIP-AVPSAG-----YNPPQMSLPVMAQPIFNFPVYVG 198
QY 119 PLPGGVPRMLITLITGVKPNANRIALDFQRGN--DVAHF--PRENNRRVIVCNTKLD 175
DB 199 TLGGGLTARRITIIKGYVLPTAKNLIINFKVGSTGDIAPFHMNPRIGD---CVRANSYMN 254
QY 176 NNMGRERQSVF-PFESGKPEKIVLVEPDHFKAIVNDAH-QYNHRVKLINEISKLGIS 233
DB 255 GSWSEBKRKIPYNPFAGQGFDDLSIRCGTDRFKVFANGQHLDFSHRFQAFQVDMLEIK 314
QY 234 GDIDLTSASYTMI 246
DB 315 GDITL---SYVQI 324

```

## RESULT 21

```

US-08-946-914-8
Sequence 8, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 200 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-8

```

```

CURRENT APPLICATION NUMBER: US/09/131,648
CURRENT FILING DATE: 1998-08-10
NUMBER OF SEQ. ID NOS: 5
SOFTWARE: PERL Program
SEQ ID NO 5
LENGTH: 316
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: 91932712
US-09-131-648-5

Query Match          16.5%; Score 223.5; DB 3; Length 316;
Best Local Similarity 39.3%; Pred. No. 6.7e-13;
Matches 55; Conservative 22; Mismatches 51; Indels 11; Gaps 5

QY      113 IVEYNIPLPGGVYPRMLITLLGVKPNANRIALDPORGN-----DYAFHF-PRFNNENNR 165
       15 VIFVFGITPQDLPQGLTIVIRGHVPSADRPQVDLQNGSSVYKPRADVAEHNPRPKRAG- 73
QY      166 RVLYCNTRKLDNMNREERQSFPEPSCKPEIQLVLPEDHPKVAVNDAH-LQYNRRVKKL 224
       74 -CIVCNTRLINEKREBEITYDTPEKREKFEIVNLKDKQVAVNGKHTLLYGRIGP- 131
QY      225 NEISKIGISGIDIDTSASYT 244
       132 EKIDTLGTYGKVNTHSIGFS 151
Db
Db
Db

RESULT 24
US-08-946-914-6
; Sequence 6, Application US/08946914
; Patent No. 6027916
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,914
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,093
; FILING DATE: 09-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SCW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-946-914-6

```



```

;
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 107880
; FILING DATE: 05-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3527
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-647-960-2

Query Match 16.1%; Score 218.5; DB 2; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VPYNLPGLGVVPRMLITITIGTVKPNANRIALDPQRGN-----DVAPEF-PRFENNRR 166
DB 16 IPYVSTIEQLKPGSLIYIRGHVPKDSERFQVDFQHGSLSKRAVDVAFHFNPRFRSN-- 73
QY 167 VIVCNTKLDNNWGEROSVFPFESGKPEKIQVLYEPDHEKAVANDAH-LQYNERVKLN 225
DB 74 CIVCNTLTNEKMGWEIETHDMFPFKESFEIYIVWLKNKFVAVANGKHILLVYHRINP-E 132
QY 226 EISKLGISGDIDLTSASY 243
DB 133 KIDTLGIKGVNIHSIGF 150

RESULT 28
US-08-946-914-15
; Sequence 15, Application US/08946914
; Patent No. 6027916
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,914
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,093
; FILING DATE: 09-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 15:
```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-946-914-15

Query Match 16.1%; Score 218.5; DB 3; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;

QY 114 VPYNLPGLGVVPRMLITITIGTVKPNANRIALDPQRGN-----DVAPEF-PRFENNRR 166
DB 16 IPYVSTIEQLKPGSLIYIRGHVPKDSERFQVDFQHGSLSKRAVDVAFHFNPRFRSN-- 73
QY 167 VIVCNTKLDNNWGEROSVFPFESGKPEKIQVLYEPDHEKAVANDAH-LQYNERVKLN 225
DB 74 CIVCNTLTNEKMGWEIETHDMFPFKESFEIYIVWLKNKFVAVANGKHILLVYHRINP-E 132
QY 226 EISKLGISGDIDLTSASY 243
DB 133 KIDTLGIKGVNIHSIGF 150

RESULT 29
US-08-946-914-17
; Sequence 17, Application US/08946914
; Patent No. 6027916
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,914
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,093
; FILING DATE: 09-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-946-914-17

Query Match 16.1%; Score 218.5; DB 3; Length 316;
Best Local Similarity 39.9%; Pred. No. 1.9e-12;
Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5;
```



```

      Db          133 KIDTLGI FGKVNHSIGF 150

      RESULT 31
      US-09-656-450-15
      / Sequence 15, Application US/09656450
      / Patent No. 6468768
      / GENERAL INFORMATION:
      / APPLICANT: Ni, Jian
      / APPLICANT: Gentz, Reiner L.
      / TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
      / FILE REFERENCE: 1488.0560003
      / CURRENT APPLICATION NUMBER: US/09/656,450
      / PRIOR FILING DATE: 2000-09-06
      / PRIOR APPLICATION NUMBER: US 09/263,689
      / PRIOR FILING DATE: 1999-03-05
      / PRIOR APPLICATION NUMBER: US 08/946,914
      / PRIOR FILING DATE: 1997-10-09
      / PRIOR APPLICATION NUMBER: US 60/028,093
      / PRIOR FILING DATE: 1996-10-09
      / NUMBER OF SEQ ID NOS: 60
      / SOFTWARE: PatentIn version 3.0
      / SEQ ID NO 17
      / LENGTH: 316
      / TYPE: PRT
      / ORGANISM: Rat
      US-09-656-450-17

      Query Match          16.1%; Score 218.5; DB 4; Length 316;
      Best Local Similarity 39.9%; Pred. No. 1.9e-12;
      Matches 55; Conservative 18; Mismatches 54; Indels 11; Gaps 5

      QY          114 VPYNLPDLGGVPRLLITLIGLVKNANRIADDPQGN-----DVAHFH-FRENNENR 166
      Db          16 IPYSTITEQLKPGSLIVIRGVPEKDSERFOVDPHGNSLKPRADVAFHPNRPKSN-- 73
      QY          167 VIVCTTKLDNNWGREEROSEVPFEESGKEPFKIQTALVEPHEFKAAVNDAA-LQYNHRVKLN 225
      Db          74 CIVCNTLTLEKXGMWEIITHDMPEFRKESFEIYIMVLKKKHVAAGSKILLIYAHRINP-E 132
      QY          226 EISKLGISGDIDLTSASY 243
      Db          133 KIDTLGI FGKVNHSIGF 150

      RESULT 32
      US-09-656-450-17
      / Sequence 17, Application US/09656450
      / Patent No. 6468768
      / GENERAL INFORMATION:
      / APPLICANT: Ni, Jian
      / APPLICANT: Gentz, Reiner L.
      / APPLICANT: Ruben, Steven M.
      / TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
      / FILE REFERENCE: 1488.0560003
      / CURRENT APPLICATION NUMBER: US/09/656,450
      / PRIOR FILING DATE: 2000-09-06
      / PRIOR APPLICATION NUMBER: US 09/263,689
      / PRIOR FILING DATE: 1999-03-05
      / PRIOR APPLICATION NUMBER: US 08/946,914
      / PRIOR FILING DATE: 1997-10-09
      / PRIOR APPLICATION NUMBER: US 60/028,093
      / PRIOR FILING DATE: 1996-10-09
      / NUMBER OF SEQ ID NOS: 60
      / SOFTWARE: PatentIn version 3.0
      / SEQ ID NO 17
      / LENGTH: 316
      / TYPE: PRT
      / ORGANISM: Rat
      US-09-656-450-17

```





QY 213 AHQYHNRVKKLNEISKIGSIDLDTLSASY 243  
DB 120 PFYEXGRL-PLQWTHLQVGDLOLQSIINF 149

## RESULT 38

US-08-946-914-13  
Sequence 13, Application US/08946914  
Patent No. 6027916  
GENERAL INFORMATION:  
APPLICANT: NI, Jian  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
STREET: 1100 New York Ave., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946,914  
FILING DATE: Herewith  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,093  
FILING DATE: 09-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Steffe, Eric K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 136 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-946-914-13

Query Match 15.5%; Score 211; DB 3; Length 136;  
Best Local Similarity 35.3%; Pred. No. 3.3e-12;  
Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;

QY 114 VPYNLPFGVVPRLITLITGVKPNANRIALDF-----QKNDVAFHF--PRENNRRVYI 168  
DB 4 VPKKSLPEGIRPGTVLRIRGLVPPNARFHVNLGCEGQSDALAHFNPRLDTSE---V 60  
QY 169 VCNTKLDNNWGREROSVFPFESGKPKIQVLVEPDHFKAIVNDA-HIQYHNRVKKLNEI 227  
DB 61 VFNSKEGSGWGERERGPVFPQGFPEVLIASDDGFKAVVDADQYHHRRL-PLARV 119  
QY 228 SKIGSIDLDTLS 240  
DB 120 RIVEVGSDVQLDS 132

RESULT 39  
US-09-154-750A-79  
Sequence 79, Application US/09154750A  
Patent No. 6432640  
GENERAL INFORMATION:  
APPLICANT: Vogelstein, Bert

APPLICANT: Kinzler, Kenneth  
APPLICANT: Polyak, Kornelia  
TITLE OF INVENTION: p53-Induced Apoptosis  
FILE REFERENCE: 1107.75357  
CURRENT APPLICATION NUMBER: US/09/154,750A  
CURRENT FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/059,153  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/079817  
PRIOR FILING DATE: 1998-03-30  
NUMBER OF SEQ ID NOS: 93  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 79  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-154-750A-79

Query Match 15.5%; Score 211; DB 4; Length 136;  
Best Local Similarity 35.3%; Pred. No. 3.3e-12;  
Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;

QY 114 VPYNLPFGVVPRLITLITGVKPNANRIALDF-----QKNDVAFHF--PRENNRRVYI 168  
DB 4 VPKKSLPEGIRPGTVLRIRGLVPPNARFHVNLGCEGQSDALAHFNPRLDTSE---V 60  
QY 169 VCNTKLDNNWGREROSVFPFESGKPKIQVLVEPDHFKAIVNDA-HIQYHNRVKKLNEI 227  
DB 61 VFNSKEGSGWGERERGPVFPQGFPEVLIASDDGFKAVVDADQYHHRRL-PLARV 119  
QY 228 SKIGSIDLDTLS 240  
DB 120 RIVEVGSDVQLDS 132

RESULT 40  
US-09-656-450-13  
Sequence 13, Application US/09656450  
Patent No. 6468768  
GENERAL INFORMATION:  
APPLICANT: NI, Jian  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides  
FILE REFERENCE: 1488.0560003  
CURRENT APPLICATION NUMBER: US/09/656,450  
CURRENT FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: US 09/263,689  
PRIOR FILING DATE: 1999-03-05  
PRIOR APPLICATION NUMBER: US 08/946,914  
PRIOR FILING DATE: 1997-10-09  
PRIOR APPLICATION NUMBER: US 60/028,093  
PRIOR FILING DATE: 1996-10-09  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 13  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-656-450-13

Query Match 15.5%; Score 211; DB 4; Length 136;  
Best Local Similarity 35.3%; Pred. No. 3.3e-12;  
Matches 47; Conservative 28; Mismatches 48; Indels 10; Gaps 5;

QY 114 VPYNLPFGVVPRLITLITGVKPNANRIALDF-----QKNDVAFHF--PRENNRRVYI 168  
DB 4 VPKKSLPEGIRPGTVLRIRGLVPPNARFHVNLGCEGQSDALAHFNPRLDTSE---V 60  
QY 169 VCNTKLDNNWGREROSVFPFESGKPKIQVLVEPDHFKAIVNDA-HIQYHNRVKKLNEI 227  
DB 61 VFNSKEGSGWGERERGPVFPQGFPEVLIASDDGFKAVVDADQYHHRRL-PLARV 119

OY 228 SKLIGSDIDLTS 240  
DB 120 RLVWGVGDVQLDS 132

## RESULT 41

US-09-557-170A-4  
; Sequence 4, Application US/09557170A  
; Patent No. 6605699  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: Galectin 11  
; FILE REFERENCE: PR354P2  
; CURRENT APPLICATION NUMBER: US/09/557,170A  
; CURRENT FILING DATE: 2000-04-21  
; PRIOR APPLICATION NUMBER: 09/109,864  
; PRIOR FILING DATE: 1998-06-06  
; PRIOR APPLICATION NUMBER: 09/010,146  
; PRIOR FILING DATE: 1998-01-21  
; PRIOR APPLICATION NUMBER: 60/034,205  
; PRIOR FILING DATE: 1997-01-21  
; PRIOR APPLICATION NUMBER: 60/034,204  
; PRIOR FILING DATE: 1997-01-21  
; PRIOR APPLICATION NUMBER: 60/169,932  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/130,390  
; PRIOR FILING DATE: 1999-04-21  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 318  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-557-170A-4

Query Match 15.3%; Score 207; DB 4; Length 318;  
Best Local Similarity 38.3%; Pred. No. 2.3e-11;  
Matches 54; Conservative 23; Mismatches 52; Indels 12; Gaps 6;

OY 113 TVPYNLPPGVPRMLTITIGVFNANRIALDFORGN-----DVAHF-PEFENNR 165  
DB 16 VIPFVGTIPDLDGTILIVIRGHPSDADRFOVDLQNGSSMKPRADVAFHNPFFKAG- 74  
OY 166 RVTYCNKTLDNMGRERQSVFPE-SGKPKIOLIVEPDHFKAANDAH-LQNHRYKK 223  
DB 75 -CIVCNLILNKKMGREBELTYDTPFOKEKSFELIYIMLKAFQVAVNGKHTLTYGHRIGP 133  
OY 224 LNEISKIGSDIDLTSASYT 244  
DB 134 -EKIDTIGTYKVNIHSGFS 153

## RESULT 42

US-09-219-849-3  
; Sequence 3, Application US/09219849  
; Patent No. 6150081  
; GENERAL INFORMATION:  
; APPLICANT: VAN HEERDE, GEORGE V.  
; APPLICANT: VAN RIJN, ALEXIS C.  
; APPLICANT: BOWMSTRA, JAN B.  
; APPLICANT: DE WOLF, FREDERIK A.  
; APPLICANT: MOORBOEK, ANDREAS  
; APPLICANT: WERTEN, MARC W.T.  
; APPLICANT: WIND, RICHEL D.  
; APPLICANT: VAN DEN BOSCH, TANJA J.  
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN  
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE  
; TITLE OF INVENTION: PREPARATION THEREOF  
; FILE REFERENCE: 2728-2  
; CURRENT APPLICATION NUMBER: US/09/219,849  
; CURRENT FILING DATE: 1998-12-23  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 3  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative  
US-09-219-849-3

Query Match 14.8%; Score 200.5; DB 3; Length 504;  
Best Local Similarity 49.2%; Pred. No. 1.6e-10;  
Matches 62; Conservative 1; Mismatches 50; Indels 13; Gaps 9;

OY 8 HDALSGSNPNPOGPGWAGNQAGAGVPG-ASYPGVPGQA-PGAYPGQA-PPGA--Y 62  
DB 12 HGPAGPKAPGAPGAPGAPG--PACPPGAPGAPGAPGAPGAPGAPGAPGAPGAP 68  
OY 63 HGAPGAYGAPAPGVYPPGPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 120  
DB 69 AGPPGA-PGAPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 125  
OY 121 PGVVP 126  
DB 126 PGAPGP 131

## RESULT 43

US-08-642-255-52  
; Sequence 52, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,255  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROWLAND, Bertam I.  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A5556-3/BIR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299 FHT UR  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 561 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-642-255-52

Query Match 14.8%; Score 200.5; DB 1; Length 561;  
Best Local Similarity 49.2%; Pred. No. 1.9e-10;  
Matches 62; Conservative 1; Mismatches 50; Indels 13; Gaps 9;



ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bertram I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A55556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 777 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-642-255-53

Query Match 14.7%; Score 199.5; DB 1; Length 777;  
Best Local Similarity 49.2%; Pred. No. 3,4e-10;  
Matches 62; Conservative 1; Mismatches 50; Indels 13; Gaps 9;

QY 8 HDALSGSNPNPQGMWPGAMGNOPAGAGVPG-ASYPGVPGA-PGAYVPGA-PPGA--Y 62  
DB 45 HGPAGPKAPGPPAGPPAG-PAQPPGAPGPPAGPPAGPPGA-PGPPAGPPAGP 101  
QY 63 HGAPAYGAPAPGVYPPGPPSGPAGVPSGQSPAPGAVATGPGYGA--PAGPLIVPNLPL 120  
DB 102 AGPPGA-PGPPAGPPAGPPAGPPAGPPAGPPAGP--PAGPPAGPPAGPPAGPPAGP 158  
QY 121 PGGVVP 126  
DB 159 PGAGP 164

RESULT 47  
US-08-642-255-51  
Sequence 51, Application US/08642255  
Patent No. 5,773,249

GENERAL INFORMATION:  
APPLICANT: CAPPELO, Joseph  
APPLICANT: FERRARI, Franco A.  
TITLE OF INVENTION: High Molecular Weight Collagen-like  
TITLE OF INVENTION: Protein Polymers  
NUMBER OF SEQUENCES: 135  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,255  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bertram I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A55556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-642-255-51

Query Match 14.6%; Score 197.5; DB 1; Length 234;  
Best Local Similarity 51.3%; Pred. No. 1,2e-10;  
Matches 61; Conservative 1; Mismatches 44; Indels 13; Gaps 9;

QY 15 GNPNGWPGAMGNOPAGAGVPG-ASYPGVPGA-PGAYVPGA-PPGA--YHGAPEAY 69  
DB 1 GAGPPAGPPAGP--PAGPPAGPPAGPPAGPPAGPPGA-PGPPAGPPAGPPAGP 56  
QY 70 PGAPAGVYPPGPPSGPAGVPSGQSPAPGAVATGPGYGA--PAGPLIVPNLPLPGVVP 126  
DB 57 PGPPAGPPAGPPAGPPAGPPAGPPAGP--PAGPPAGPPAGPPAGPPAGPPAGP 113

RESULT 48  
US-08-526-136-13  
Sequence 13, Application US/08526136  
Patent No. 6,107,089

GENERAL INFORMATION:  
APPLICANT: Towle, Christine A. et al.  
TITLE OF INVENTION: ANNEXIN XI  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: WordPerfect (Version 5.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/526,136  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,036  
FILING DATE:  
APPLICATION NUMBER: 07/837,775  
FILING DATE: February 13, 1992  
APPLICATION NUMBER: 07/764,465  
FILING DATE: September 23, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/099001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 466  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: N/A  
US-08-526-136-13

Query Match 14.5%; Score 196.5; DB 3; Length 466;  
Best Local Similarity 38.8%; Pred. No. 3,4e-10;  
Matches 57; Conservative 6; Mismatches 45; Indels 39; Gaps 11;

QY 17 PNPQGMWPGAMGNOPAGAGVPGASYPGAPPGAYVPGA-PGAYVPGA-PGAYVPGA-PGAP 76  
DB 33 PYPSPG-----PMGGAYPVPSSGYPG---AGGTPA---PGY-PAFGYFGAQP 79  
QY 77 ---VYPPGPPG-----PGAYPSSGQSPAGP-AVATGPGY-----GAPAGPLIVPNLPL 120

Db 80 GAPSYGVPPGGFVPGAGFSGYPPPSQSYSGGAPVPLPGFPGQMPQY---- 135  
QY 121 PGK--VPRMLTITL-----GVKPNAN 141  
Db 136 PGGPTPSQPAIVQVQGITRPAAN 162

## RESULT 49

US-08-175-155-69  
; Sequence 69, Application US/08175.155  
; Patent No. 5641648

## GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A.  
APPLICANT: Cappello, Joseph  
APPLICANT: Crisman, John W.  
APPLICANT: Dorman, Mary A.  
TITLE OF INVENTION: Methods for Preparing Synthetic  
TITLE OF INVENTION: Repetitive DNA  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/175,155  
FILING DATE: 29-DEC-1993

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Rowland, Bettiam I.  
REGISTRATION/DOCKET NUMBER: 20015  
REFERENCE/DOCKET NUMBER: A-55186-5/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249

## INFORMATION FOR SEQ. ID NO: 69:

SEQUENCE CHARACTERISTICS:  
LENGTH: 417 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-175-155-69

Query Match 14.4%; Score 195.5; DB 1; Length 417;  
Best Local Similarity 40.1%; Pred. No. 3.7e-10;  
Matches 55; Conservative 8; Mismatches 55; Indels 19; Gaps 7;

QY 9 DALSSGNNPQGWGAWGNQ----PAGAGYVGAHYPPG-QAPPGA--YFGQAPPA 61  
Db 31 DPMGAGTGPQGLPGSPGAPGTGPQGLPGSPGA--PCTPGQGLPGSPGAPGTGPQ 88  
QY 62 YGAPGAYGAPAPGYPPSPGAPGAYPSGQPSAFGAYAT-GPYGAPA-----GPI 112  
Db 89 LRGSPGA-PGTGPQGLPGSPGAPGTGPQGLPGSPGAPGTGPQGLPGSPGAPGTGPQ 147  
QY 113 IVPYNLPLPGVVRML 129  
Db 148 GLPGSPGAPGTGPQGL 164

RESULT 50  
US-08-477-509B-104  
; Sequence 104, Application US/08477509B  
; Patent No. 5770697

## GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A  
APPLICANT: Cappello, Joseph  
APPLICANT: Crisman, John W  
APPLICANT: Dorman, Mary A  
TITLE OF INVENTION: No. 5770697e1 Peptides Comprising Repetitive  
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same  
NUMBER OF SEQUENCES: 112  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94111

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,509B  
FILING DATE: 07-JUN-1995

## CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/175,155  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,049  
FILING DATE: 22-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987

## APPLICATION NUMBER: US 06/927,258

## FILING DATE: 04-NOV-1986

## ATTORNEY/AGENT INFORMATION:

NAME: Treacartin, Richard F.  
REGISTRATION/DOCKET NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249

## INFORMATION FOR SEQ. ID NO: 104:

SEQUENCE CHARACTERISTICS:  
LENGTH: 417 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-509B-104

Query Match 14.4%; Score 195.5; DB 1; Length 417;  
Best Local Similarity 40.1%; Pred. No. 3.7e-10;  
Matches 55; Conservative 8; Mismatches 55; Indels 19; Gaps 7;

QY 9 DALSSGNNPQGWGAWGNQ----PAGAGYVGAHYPPG-QAPPGA--YFGQAPPA 61  
Db 31 DPMGAGTGPQGLPGSPGAPGTGPQGLPGSPGA--PCTPGQGLPGSPGAPGTGPQ 88  
QY 62 YGAPGAYGAPAPGYPPSPGAPGAYPSGQPSAFGAYAT-GPYGAPA-----GPI 112  
Db 89 LRGSPGA-PGTGPQGLPGSPGAPGTGPQGLPGSPGAPGTGPQGLPGSPGAPGTGPQ 147  
QY 113 IVPYNLPLPGVVRML 129  
Db 148 GLPGSPGAPGTGPQGL 164

Search completed: August 23, 2004, 14:39:25  
Job time : 24 secs